

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 16:43:35 ; Search time 9219 Seconds  
(without alignments)  
11326.702 Million cell updates/sec

Title: US-09-915-706a-1

Perfect score: 3588

Sequence: 1 gcgcactatcgtatgatgcgttaa 3588

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_srs:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pin:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description        |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1          | 82.8  | 2.3         | 190050 | 1  | AL646080 | AL646080 Ralstonia |
| 2          | 80    | 2.2         | 33154  | 1  | AF361470 | AF361470 Rhizobium |
| 3          | 79.4  | 2.2         | 14867  | 3  | AE001398 | AE001398 Plasmid   |
| 4          | 76.2  | 2.1         | 179902 | 2  | AP003630 | AP003630 Oryza sat |
| 5          | 75.4  | 2.1         | 143409 | 9  | AL162497 | AL162497 Human DNA |
| 6          | 73    | 2.0         | 83703  | 5  | AL732635 | AL732635 Zebrafish |
| 7          | 72.8  | 2.0         | 11422  | 6  | AX345121 | AX345121 Sequence  |
| 8          | 72.8  | 2.0         | 11422  | 6  | AX348323 | AX348323 Sequence  |
| 9          | 72.8  | 2.0         | 224149 | 2  | AC119799 | AC119799 Pan trogl |
| 10         | 72.8  | 2.0         | 349980 | 6  | AX344564 | AX344564 Sequence  |
| 11         | 72.6  | 2.0         | 5727   | 1  | AF037441 | AF037441 Edwardsie |
| 12         | 71.4  | 2.0         | 61020  | 6  | AX251546 | AX251546 Sequence  |
| 13         | 71.4  | 2.0         | 131033 | 5  | AL591370 | AL591370 Zebrafish |
| 14         | 70.8  | 2.0         | 13568  | 1  | AE009363 | AE009363 Agrobacte |
| 15         | 70.8  | 2.0         | 19961  | 1  | AE008251 | AE008251 Agrobacte |
| 16         | 70.6  | 2.0         | 10029  | 1  | AE013970 | AE013970 Yersinia  |
| 17         | 70.6  | 2.0         | 72243  | 9  | AL731858 | AL731858 Human DNA |
| 18         | 70.6  | 2.0         | 201050 | 9  | AJ414143 | AJ414143 Yersinia  |
| 19         | 70.2  | 2.0         | 101034 | 9  | AC096550 | AC096550 Homo sapi |
| 20         | 69.8  | 1.9         | 86827  | 3  | PFMA13P5 | PFMA13P5 Plasmid   |
| 21         | 69.4  | 1.9         | 8093   | 6  | AX251190 | AX251190 Sequence  |
| 22         | 69.4  | 1.9         | 8093   | 6  | AX281167 | AX281167 Sequence  |
| 23         | 69.4  | 1.9         | 8093   | 6  | AX347411 | AX347411 Sequence  |
| 24         | 69.4  | 1.9         | 8093   | 6  | AX349132 | AX349132 Sequence  |
| 25         | 69.4  | 1.9         | 8093   | 6  | AX356412 | AX356412 Sequence  |
| 26         | 69.4  | 1.9         | 131682 | 9  | AL672277 | AL672277 Human DNA |
| 27         | 69.2  | 1.9         | 1141   | 6  | AX083744 | AX083744 Sequence  |
| 28         | 69    | 1.9         | 5504   | 6  | AX252009 | AX252009 Sequence  |
| 29         | 69    | 1.9         | 5504   | 6  | AX348782 | AX348782 Sequence  |
| 30         | 69    | 1.9         | 105574 | 2  | AC013308 | AC013308 Homo sapi |
| 31         | 69    | 1.9         | 125097 | 9  | AL691415 | AL691415 Human DNA |
| 32         | 69    | 1.9         | 161442 | 2  | AL683875 | AL683875 Homo sapi |
| 33         | 69    | 1.9         | 162377 | 2  | AL683870 | AL683870 Homo sapi |
| 34         | 68.8  | 1.9         | 191722 | 9  | AL16360  | AL16360 Homo sapi  |
| 35         | 68.6  | 1.9         | 837    | 6  | A85629   | A85629 Sequence 28 |
| 36         | 68.6  | 1.9         | 837    | 6  | AR155122 | AR155122 Sequence  |
| 37         | 68.6  | 1.9         | 837    | 6  | E65647   | E65647 Genome DNA  |
| 38         | 68.4  | 1.9         | 840    | 8  | CNS0180K | AL110675 Botrytis  |
| 39         | 68.2  | 1.9         | 51014  | 2  | AC011856 | AC011856 Homo sapi |
| 40         | 68.2  | 1.9         | 83391  | 6  | AX458578 | AX458578 Sequence  |
| 41         | 68    | 1.9         | 6306   | 6  | AX278061 | AX278061 Sequence  |
| 42         | 68    | 1.9         | 6306   | 6  | AX323844 | AX323844 Sequence  |
| 43         | 68    | 1.9         | 13131  | 6  | AX356424 | AX356424 Sequence  |
| 44         | 67.2  | 1.9         | 60604  | 2  | AC023466 | AC023466 Homo sapi |
| 45         | 66.8  | 1.9         | 99003  | 2  | AL390756 | AL390756 Homo sapi |

#### ALIGNMENTS

RESULT 1  
AL646080  
LOCUS  
DEFINITION  
AL646080 190050 bp DNA linear BCT 07-DEC-2001  
segment 5/11.  
ACCESSION  
AL646080 AL646053  
VERSION  
AL646080.1 GI:17431113  
KEYWORDS  
SOURCE  
ORGANISM  
Ralstonia solanacearum.  
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
Ralstonia.  
REFERENCE  
1 (bases 1 to 190050)  
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,  
Arlat,M., Billault,A., Brotier,P., Camus,J.C., Catolico,L.,

Pred. No. is the number of results predicted by chance to have a

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL  
COMMENT  
FEATURES  
source  
Chandler, M., Choisine, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavié, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weissensbach, J. and Boucher, C. A.  
Genome sequence of the plant pathogen *Ralstonia solanacearum*  
Unpublished  
2 (bases 1 to 190050)  
Boucher, C. A.  
Direct Submission  
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
Bp27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMG CNRS  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5106, 91057 Evry Cedex, France,  
Laboratoire de Biometrie et Intelligence Artificielle INRA, Bp27,  
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire  
INRA, Bp27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.  
Location/Qualifiers  
1. 190050  
/organism="Ralstonia solanacearum"  
/strain="GM11000"  
/db\_xref="taxon:305"  
/plasmid="megaplasmid"  
51. 362  
/gene="Rsp0643"  
/note="RS05579"  
51. 362  
/gene="Rsp0643"  
/function="miscellaneous; unknown"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="HYPOTHETICAL PROTEIN"  
/protein\_id="CADI7794.1"  
/db\_xref="GI:17431114"  
/translation="MKTSPSPNPAAGPASDPPEAGDGVAFDQLHEYNHTTIRPR  
RIGGGLMRAFAVSLDSRALVAVTEVGAFFPICARDAVDAARSFIDARLKACRP  
A"  
397. 555  
/gene="Rsp0644"  
/note="RS05578"  
397. 555  
/gene="Rsp0644"  
/function="miscellaneous; unknown"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"  
/protein\_id="CADI7795.1"  
/db\_xref="GI:17431115"  
/translation="MERQQLFAIAIVIAAVSAAYSLARDLNDERDFPGQALVHGH  
ATASAAGR"  
complement(575. 1135)  
/gene="Rsp0645"  
/note="RS05577"  
complement(575. 1135)  
/gene="Rsp0645"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon\_usage

gene  
CDS  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CADI7796.1"  
/db\_xref="GI:17431116"  
/translation="MKRISTFPPDVMKKVAIKERSPAAPGQALSLDRTDRAL  
KMLORDASISNVALADKVSPPACRLRVERLKEICIRGVALLLEPKALDLTLVMI  
GYLDRSTPESPAFDEKAVOKVGCLECHVVTGEPFVFMVPTRTDNDYSRLAEQQL  
YLGVGRQIRTFMILKYLTQLPPIA"  
1244. 2260  
/gene="Rsp0646"  
/note="RS05576"  
1244. 2260  
/gene="Rsp0646"  
/EC\_number="4.1.99.4"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE  
DEAMINASE (ACC DEAMINASE) PROTEIN"  
/protein\_id="CADI7797.1"  
/db\_xref="GI:17431117"  
/translation="MMLNKHPRHPLTFGPPPIOLRLRSLAHLGKVELYAKREDNSG  
LAFGKNKTRKLEVLVPEVLGAGYDTLVISIGISNOTRQYAAVAHIGLGVJOEWM  
VAVADAYDVRVGNIELSRILGADVRLDAAGFDIGIRSPWQAEADYRAGKPEPIPA  
GCEHPILGIGFVGFAPFEVROQAEGERRDYIVVCSYVSGTAGMVGVAAGRGRD  
VIGIDSAKPEQTRTQRLTARTAKIVELGRITTEDDVLDTIRGPEVPELVNCSFL  
EATRLCARQEGMLTDPVEYEGKSMHMDTRVGRGGEPEGSGVYLTAHLGVPALNVSFL  
FRNG"  
complement(2336. 3508)  
/gene="Rsp0647"  
/note="RS05575"  
complement(2336. 3508)  
/gene="Rsp0647"  
/EC\_number="4.2.1.17"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"  
/protein\_id="CADI7798.1"  
/db\_xref="GI:17431118"  
/translation="MSALPKPAEPAAPAAAEVLEHVNVRAILTLNRPALNALSHQM  
VROLSALACQDDPRLAAVYLKAGAGCGFCAGGYALYRLRACQAHGPDMLQFF  
VDYRILDYALHRRPKPVYALDGLMGCGGGLGQARLRYTERSRAMPEPRIGFLP  
DVGATHTLGAMPPEMALXIGLTSTGSGAPALCLGADNVAPAEMLASPEDRLLRLSH  
AGPDGIDLMOLARVFEPCNIVPAAIARTTPVIVRFDRRSTIDRIYATLRSLERD  
PAREARQWLQATLDAMATSHPTMLHYTRERLLNGRQTLTAECFRMEIGTVARAIEGD  
FCEGVRHAILVDKCHKHRRGMAPATLAGYRPERVRHFLSSPWRGAAGPLAHLGAPOAA"  
complement(3505. 4305)  
/gene="Rsp0648"  
/note="RS05574"  
complement(3505. 4305)  
/gene="Rsp0648"  
/EC\_number="4.2.1.17"  
/function="small molecule metabolism; degradation; carbon  
compounds"  
/note="Product confidence : probable  
Gene name confidence : hypothetical



SLKFSMSADPEFGAAVDDVPGIKALLNARNQJLRDMSKADRSSELERLLEDIIQONKAD  
LSOLVLAIGCGKADAKDALN"  
2533..4014 /gene="impc"  
2533..4014 /gene="impc"  
/function="involved in temperature-dependent protein  
secretion"  
/note="hypothetical; similar to Edwardsiella ictaluri  
hypothetical protein Eip55"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="impc"  
/protein\_id="AAI17801.1"  
/db\_xref="GI:16326479"  
/translation="MTAEAFDTKEVDGAPDENLSKVAAARTQTEPRDRAOOLRLT  
TDQALGVYDUTLITLNLIAETDKISQOLAIMOSPETTKLEGAMRGILNYLTK  
NSETSANKIRLYNLSKRDVADKLAFAFDSDRLKSYEDEGFTGGPGMALIGD  
YEFDSFEDVODLQGVSMIAAFAFAFISAASPRMGFEDEFRLAPRDLKTEFVE  
YAKWMSLSDSDSRVFTVLAMPVILRMYPGPTSPTEEFNFDETGSGKSLSHGSC  
WMNAYVWGTRLTEAFKMGCTAIRKAPENGKVEGLPMHIFSSDGDLDLKKPTGEG  
ITDRDAALTKGKGLPLCHYKNTDVAFPGAGTARHKPKYDREPTANAAVASRLPYM  
MATSREDAIYKVMGDKIGSPMEADGCEAMLRWISNYNANDAGEESAKTPIRLRA  
KVYVQETIGRPGCAIYAVAMMRPWLQNEELTSLRWARLPSKN"  
4081..5469 /gene="impc"  
4081..5469 /gene="impc"  
/function="involved in temperature-dependent protein  
secretion"  
/note="hypothetical; similar to Impc and to Edwardsiella  
ictaluri hypothetical protein Eip55; possible duplication  
of Impc"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="impc"  
/protein\_id="AAI17802.1"  
/db\_xref="GI:16326480"  
/translation="GGSSSEMRDRDOLITLIDEALNKQVNAIHHHPQAMARNGR  
LAMLYRSESEVYKILGVWSELTETMERASDPDOSHPELIYSREMGCGEPL  
GLVYDYLSPADPAGGDPVSTLSQIGVAAAACFPVGAAPLAIGLEDPEHLEIRVN  
DFSWLANTATIRRMGLRARREDRELGIYAPILKMSPLTPHSRNNNDGPPRESLAG  
DGSLSQAVEHIEFATVAVIIRNIDSGMFADIRGVTQDAVDGGLSTAEPLPYDFGES  
NGLSKQAVEHIEFATVAVIIRNIDSGMFADIRGVTQDAVDGGLSTAEPLPYDFGES  
ONARLAPMLQVYLCSRSFHYKVMRDEITGOLADISIERKADMLTATVLTGNDQ  
IALRPLPLRSLSGICVAELPKGRGSGTSCVIRLQPHQDLDDVATISFHLIATATGQIL  
SRPVATPRFMS"  
5469..6296 /gene="impc"  
5469..6296 /gene="impc"  
5469..6296 /gene="impc"  
/function="involved in temperature-dependent protein  
secretion"  
/note="hypothetical"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="impc"  
/protein\_id="AAI17803.1"  
/db\_xref="GI:16326481"  
/translation="MTLSKSIQSISLSDNALDIDAIEVKAHLKVPSPDOEARHLIYIDL  
VLAVGQADNCSIAATISDPATMGFALLRNLRAMAAMDAMFTGGAIDEPFGGSE  
LDKLAVRGINAHROSNADDEARTLDALTELKGERPLINNGKASDPRDLDDLRPHALE  
VIMTGGGLWIDFAKIAAISLPIAPRDLARRALSLIDGAAASVLLPAAVHGTEK  
DETURLGRETETIEETGTISGRGRCYLAGDEWSPHDQSLIIVPAAAAGROVAVHG

| Query Match | Best Local Similarity | Matches       | 183;    | Conservative | 54.3%;  | Pred.  | No. 1.6e-06; | Mismatches | 151;   | Indels  | 3;      | Gaps  | 1;   |
|-------------|-----------------------|---------------|---------|--------------|---------|--------|--------------|------------|--------|---------|---------|-------|------|
| QY          | 3232                  | AACATAAGCGCGT | TAGT    | AAGA         | CGCGT   | GTC    | CAGCAT       | CCTAT      | ATG    | ACGTT   | GAACGA  | TGGCG | 3311 |
| Db          | 1897                  | AACGTGAACGGG  | GTTCG   | CAAC         | CGCGAGT | CCAT   | CATCAAT      | AATACGA    | AACTGG | AAACCG  | GAAGCG  | 1956  |      |
| QY          | 3312                  | CCGTAAGACGAA  | AGAGCG  | CGCGCT       | TGGT    | TGGT   | GGCGT        | CAAT       | GGCG   | CAATTT  | CGAGACA | 3371  |      |
| Db          | 1957                  | CGATGGGTGAAG  | AACGCG  | CGCTT        | GTGTC   | GTGTC  | GGCGT        | CGCG       | CACTT  | CTCCG   | GAATTC  | 2016  |      |
| QY          | 3372                  | AACCAATCAGAA  | AAAGAT  | TGATT        | TAGAA   | AGCGAG | AGAT         | CTCAC      | GGTAT  | CGATAA  | GACA    | 3431  |      |
| Db          | 2017                  | CCACGACCGCG   | TCAA--- | GGCTT        | GGCGG   | CGGCA  | AGTT         | CGT        | GAGGT  | CGACCG  | GACA    | 2073  |      |
| QY          | 3432                  | ACTTCGATACAG  | TGATGG  | GGGCA        | AAATTC  | ACCGCG | CTTTT        | CGTAC      | GAAGTT | GTGATAA | CAAGC   | 3491  |      |

| Db  | 2074   | ACTTGCAGAGTCATATGCGCCGATGACGCCGGGCGCTACCCCTGTGCGGTTCGAGAACATC             | 2133  |
|---|--|---|---|
| Qy  | 3492   | TTGCTATGATGATGATGACCATGTTTGAAGTGAACCTTGACCTCCGTTTCGATGAAGAATTC            | 3551  |
| Db  | 2134   | TTCGAGAAAGAACGGAACAACGCTTCCCGCTTGTTCGTAATTCGAAAGCAATGCCGATTCG             | 2193  |
| Oy  | 3552   | ACCGAGAACCTTGTGATGATNAATTTAGCGCGCTTAA                                     | 3588  |
| Db  | 2194   | AACCGGCGCGGTCTGATCAGGTGCCGGGCGTAA   | 2230  |
| RESULT 3  | LOCUS  | DEFINITION  | ACCESSION   |
| AE001398  | AE001398   | 14867 bp DNA linear INV 16-Apr-2002                                       | Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.                                  |
| KEYWORDS  | SOURCE   | ORGANISM  | REFERENCE   |
| AE001398.1 GI:3845197                                     | Plasmodium falciparum 3D7.   | Plasmodium falciparum 3D7   | Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.   |
| REFERENCE   | AUTHORS  | TITLE   | JOURNAL   |
| 1 (bases 1 to 14867)                                      | Gardner M.J., Tetteelin H., Carucci D.J., Cummings L.M., Aravind L., Koonin E.V., Shalimov S., Mason T., Yu K., Fujii C., Pederson J., Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Berta M., Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O., Fraser C.M., Adams M.D., Venter J.C. and Hoffman S.L. | Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum | Science 282 (5391), 1126-1132 (1998)  |
| REFERENCE   | AUTHORS  | TITLE   | JOURNAL   |
| 2 (bases 1 to 14867)                                      | Gardner M.J.   | Direct Submission   | Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA |
| FEATURES  | source   | gene  | CDS   |
| 1. 14867  | /organism="Plasmodium falciparum 3D7"  |   |   |
| /strain="3D7"   | /db_xref="taxon:36329"   |   |   |
| /chromosome="2"   | complement(1570..2424)   |   |   |
| /gene="PFB0490C"  | complement(1570..2424)   |   |   |
| /gene="PFB0490C"  | /note="predicted by GlimmerM"  |   |   |
| /codon_start=1  | /product="hypothetical protein"  |   |   |
| /protein_id="AAC71887.1"                                  | /db_xref="GI:3845198"  |   |   |
| /translation="MKKEKKEINDYSLCPEDVDYDRKKSGKSLKSSSTKKSIDYK   |  |   |   |
| SLSPKSRDHSKGSSTGRGRSGFTINKSSFKPYNNRLINNNYNGRGNHNR         |  |   |   |
| DNEGFGSGSRYFDMRKGSPFKRFTSNRKSSTVKYSGNSKMGKSPNKPTSR        |  |   |   |
| TWTYKRLNNYKTVSAPVKPKFNLLNISLYRKNRPALPKRSKPYGTTKSSVPRKRIKK |  |   |   |
| GLKSGSLKSTKRSKTSKSGSKPKLKKYLSKIKITVSLNKIPLKEQKNEVNLPESL   |  |   |   |
| NNATTKNN"   |  |   |   |
| 11241..14606  | /gene="PFB0495W"   |   |   |
| 11241..14606  | /gene="PFB0495W"   |   |   |
| /note="predicted by GlimmerM"                             | /codon_start=1   |   |   |
| /product="hypothetical protein"                           | /protein_id="AAC71888.1"   |   |   |
| /db_xref="GI:3845199"                                     |  |   |   |
| /translation="MFALKKNTVREGFVNICFSYLKLYLKSNFVTVNLVNETNEKR  |  |   |   |
| TNKKYIKSSKAQSLDFGLNTHDKLILFKNLPKKCAKAYEISAKYVKKYLDEKKG    |  |   |   |
| ENYSLCDIIOSSYIPDELKTPFDYNYFIEVKRIDKNNVLKINIEYIPKKDITFHR   |  |   |   |
| EILCKITKNKHSYTHENKNGELHFLIYFRRWNNQNLILFTYNYFVDFHLYFN      |  |   |   |
| HEIYKLLFENKYLNNNSNIPFNKNLIQEMEFNLYFREIKNEKNYIKNKKKILYKK   |  |   |   |

```

CEAFHFHNDHIDNEKLENTLRLVYDVSIIIDIDNNMCLNCLNNLNNENIEYISKL
NEFOTLKKGRYDMDMTYIKLKEVIKAFVNHLCOKTINSLKTFPSDIDYTLNLSNKK
FILNKLIDKNFIELEECLELILNKLKRVHNSQSLCISLISLETNYNLRRNYYIVNKK
LEFNIDKFIENFLEGRKIKETENEAVALIHHNDQTSNKNENIMIDIILOKRIKEY
LEFYMENKSYHLCEFLKIDRSIDKISLNTPEVKINEVNSDEYLLFPNISCIYVFLY
NRRNVRKKDYIYIILNDSPVYVYIKNNDRPKKKRKKFLELSSMKKILCKNLISVN
RYIKHHEEDNEFDOKDOYVSUFLFNLPFDKILHPIYINLMCHVKYKTYNRECKNL
INEDIISLLILJCSKFOYFIENNSNDYCRKELIHLKYNLIIDILKYNLWYSISID
NISKFIISNSKYTCVENNLLESLOSFEKVKTKSGKGIHMOBNLIDNNNSCE
KYEHRYIEYKKNELFLNKLIECLILNIFLYLKKRYGLYUOLOSCLINLEKIL
KILYIANNLYIEMYGVCEMLEERVLSIIKSEOMLFYNNKNEHMPKILCHISE
DDYIENSNTMYVLYFYDKLNIINSPROSNIIIRNSETRPFIDEIKERKYLNNNTLJKH
NNKLANLEKSNNSNGNISNLKDKNKNHNNVENDLIDNKNENKIDCKONGENCEN
CKDVLVDIINIFELKMEKKFLFOLYVLCNITKFKRRYVSSSLFPHMDFKIIR
DMLNKLYVCEKIKKNECAFVLYIDIVLEKFR*

```

| LOCUS                 | AP003630   | 179902 bp   | DNA | Linear | HTG 21-MAR-2002 |
|-----------------------|--|-------------|-----|--------|-----------------|
| DEFINITION            | Oryza sativa (japonica cultivar-group) chromosome 6 clone P0566A10.  |             |     |        |                 |
| ACCESSION             | AP003630   |             |     |        |                 |
| VERSION               | AP003630.1   | GI:14141691 |     |        |                 |
| KEYWORDS              | HTG; HTGS; PHASE2.   |             |     |        |                 |
| SOURCE                | Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA, clone: P0566A10.  |             |     |        |                 |
| ORGANISM              | Oryza sativa (japonica cultivar-group)   |             |     |        |                 |
| REFERENCE             | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartioideae; Oryzaceae; Oryza.  |             |     |        |                 |
| AUTHORS               | 1  |             |     |        |                 |
| TITLE                 | Sasaki, T., Matsumoto, T. and Yamamoto, K.   |             |     |        |                 |
| JOURNAL               | Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC clone: P0566A10  |             |     |        |                 |
| REFERENCE             | Published Only in Database (2001)  |             |     |        |                 |
| AUTHORS               | 2 (bases 1 to 179902)  |             |     |        |                 |
| TITLE                 | Sasaki, T., Matsumoto, T. and Yamamoto, K.   |             |     |        |                 |
| JOURNAL               | Submitted (17-MAY-2001) Takui Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kamondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan   |             |     |        |                 |
| COMMENT               | (E-mail: tsasaki@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)  |             |     |        |                 |
| NOTE:                 | NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved. |             |     |        |                 |
| FEATURES              | * NOTE: This is a 'working draft' sequence.  |             |     |        |                 |
| source                | * This sequence will be replaced by the finished sequence as soon as it is available and * the accession number will be preserved.   |             |     |        |                 |
| BASE COUNT            | 50739 a 38352 c 38680 g 51781 t 350 others   |             |     |        |                 |
| ORIGIN                |  |             |     |        |                 |
| Query Match           | 2.1%; Score 76.2; DB 2: Length 179902.   |             |     |        |                 |
| Best Local Similarity | 48.5%; Pred. No. 9.7e-06;  |             |     |        |                 |
| Matches               | 210; Conservative 0; Mismatches 223; Indels 0; Gaps 0;   |             |     |        |                 |
| OY                    | 514 GTTGGTGGCTTAATTCGTTACTGCTCATATTCAATTAATTCACATTAATAACAGTT 573   |             |     |        |                 |
| DB                    | 63539 GTTTATTTTTTTTTTTTTTTTATTTCTTTTCTTTTCTTTTCTTTTCTTTGTTATTTGTT 63480  |             |     |        |                 |
| OY                    | 574 CTAAAGAGCTGTTATTCGATGAATATTCGAAATTAATCACAATAAATTGATGCTATTAT 633  |             |     |        |                 |
| DB                    | 63479 TTTTCTTTTTTTTTTTTTTGTGATTTTTTTTTTTTTTTTAAATCTTTTTTTTTTTTTT 63420   |             |     |        |                 |
| OY                    | 634 TACTTCGCTGATTCGATCAACCTTCATGCGCTATATACATGAATATTTTCAGTTAGAC 693   |             |     |        |                 |
| DB                    | 63419 TCGTTTAAATTTTAAATTTTTTTTTTTTTTTTAATTTGTTTTTAATATTTGGTTTTTTT 63380  |             |     |        |                 |
| OY                    | 694 CTTAATTCAGGTAATTGCTCTAATTAATTAATTAATTCGATAATATGATTCGATGCTT 753   |             |     |        |                 |
| DB                    | 63359 TTTTCTTTTTTAATTTTGGTTTTTAAATTTTTTTTTTGTATTAATTAATTTTATGCT 63300  |             |     |        |                 |
| OY                    | 754 TGTGGTAAATTTTAAATGTTGTTCAATTTTAAATGACGGGACCTGTCATCATATTTT 813  |             |     |        |                 |
| DB                    | 63299 TTTTGTTCCTAAGTGTGTTTTTAATTTTCTTTTTTAATTTTTTTTTTAACTTTTTTT 63240  |             |     |        |                 |
| OY                    | 814 TTATGATGACAACATCTTTGATGAGATTAATTAAGATATTTGTAATGCAATGAGGCTTGC 873   |             |     |        |                 |
| DB                    | 63239 ATTTTTCCTAATATTTTTTTTTTTTTTTTTTTTAAATTTCCTTTTTTTTTTAAATTTTA 63180  |             |     |        |                 |

| QY         | 874   | CGCTTTTATTTATTTATTAATCATATTAATAACAAATATATGTTATGCTATTTT       | 933                        |
|------------|---|--|----------------------------|
| Db         | 63179   | TTATTTTATTTTATTTTATTTTATTTATATGATTTGATTTATTTTTCGTTATTTTATTTT | 63120                      |
| QY         | 934   | ATAGTGTCTCTTTT   | 946                        |
| Db         | 63119   | TTTTTGTCTTTT   | 63107                      |
| RESULT 5   |   |  |                            |
| LOCUS      | AL162497/c  |  |                            |
| DEFINITION | AL162497  | 143409 bp  | DNA linear PRI 06-JUN-2001 |
| ACCESSION  | AL162497  |  |                            |
| VERSION    | AL162497.20   |  | GI:14329908                |
| KEYWORDS   | HTG.  |  |                            |
| SOURCE     | human.  |  |                            |
| ORGANISM   | Homo sapiens  |  |                            |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.   |  |                            |
| AUTHORS    | 1 (bases 1 to 143409)   |  |                            |
| TITLE      | Mashreghi-Mohammadi, M.   |  |                            |
| JOURNAL    | Direct Submission   |  |                            |
| COMMENT    | Submitted (05-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humgen@esanger.ac.uk clone requests: clonerequests@sanger.ac.uk<br>On Jun 8, 2001 this sequence version replaced gi:14280409.<br>During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.<br>This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPep; Information on the WormPep database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <a href="http://www.sanger.ac.uk/HGP/Chr13">http://www.sanger.ac.uk/HGP/Chr13</a><br>Rp11-313L9 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/bacpac/home.htm">http://www.chori.org/bacpac/home.htm</a><br>VECTOR: pBAC3.6<br>IMPORTANT: This sequence is not the entire insert of clone Rp11-313L9 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.<br>The true right end of clone Rp11-313L9 is at 143409 in this sequence. The true left end of clone Rp11-40E6 is at 100074 in this sequence. The true right end of clone Rp11-35F13 is at 100 in this sequence. |  |                            |
| FEATURES   | Location/Qualifiers   |  |                            |
| SOURCE     | 1..143409   |  |                            |
|            | /organism="Homo sapiens"  |  |                            |
|            | /db_xref="taxon:9606"   |  |                            |
|            | /chromosome="13"  |  |                            |
|            | /clone="Rp11-313L9"   |  |                            |
|            | /clone_id="RPCI-11.2"   |  |                            |
|            | 1..350  |  |                            |
|            | /note="11map repeat: matches 5421..5789 of consensus"   |  |                            |
|            | 351..654  |  |                            |
|            | /note="11p1 repeat: matches 5846..6155 of consensus"  |  |                            |
|            | 655..957  |  |                            |
|            | repeat_region   |  |                            |
|            | repeat_region   |  |                            |
|            | repeat_region   |  |                            |

|                 |  |
|-----------------|--|
| repeat_L1region | /note="L1MA9 repeat: matches 5789. .6085 of consensus 2061. .2106    |
| repeat_L1region | /note="23 copies 2 mer tg 91% conserved" 2063. .2106                 |
| repeat_L1region | /note="11 copies 4 mer tggg 93% conserved" 2138. .2422               |
| repeat_L1region | /note="L1MD3 repeat: matches 6691. .7023 of consensus" 2422. .3051   |
| repeat_L1region | /note="L1MC4 repeat: matches 7168. .7837 of consensus" 3674. .4545   |
| repeat_L1region | /note="L1PA4 repeat: matches 5272. .6144 of consensus" 5701. .5910   |
| repeat_L1region | /note="7 copies 30 mer 8% conserved" 6490. .6613                     |
| repeat_L1region | /note="MR repeat: matches 92. .225 of consensus" 9599. .9648         |
| repeat_L1region | /note="25 copies 2 mer aa 94% conserved" 10251. .10559               |
| repeat_L1region | /note="A1USX repeat: matches 2. .312 of consensus" 11134. .11521     |
| repeat_L1region | /note="L1MB7 repeat: matches 5762. .6161 of consensus" 12667. .12710 |
| repeat_L1region | /note="11 copies 4 mer tggg 81% conserved" 13210. .13309             |
| repeat_L1region | /note="L2 repeat: matches 2559. .2674 of consensus" 13794. .14148    |
| repeat_L1region | /note="THIC repeat: matches 1. .371 of consensus" 14816. .15030      |
| repeat_L1region | /note="MR repeat: matches 49. .256 of consensus" 14982. .15037       |
| repeat_L1region | /note="L2 repeat: matches 2651. .2705 of consensus" 15476. .15940    |
| repeat_L1region | /note="L2 repeat: matches 1314. .1835 of consensus" 16305. .16609    |
| repeat_L1region | /note="A1USX repeat: matches 1. .300 of consensus" 16683. .16722     |
| repeat_L1region | /note="20 copies 2 mer tt 77% conserved" 16948. .17073               |
| repeat_L1region | /note="ML1C repeat: matches 348. .464 of consensus" 17074. .17360    |
| repeat_L1region | /note="A1UJo repeat: matches 13. .299 of consensus" 17361. .17661    |
| repeat_L1region | /note="ML1C repeat: matches 7. .348 of consensus" 17974. .18347      |
| repeat_L1region | /note="L1MD1 repeat: matches 5659. .6029 of consensus" 18399. .18554 |
| repeat_L1region | /note="3 copies 52 mer 75% conserved" 18400. .18549                  |
| repeat_L1region | /note="5 copies 30 mer 73% conserved" 18413. .18552                  |
| repeat_L1region | /note="5 copies 28 mer 74% conserved" 18565. .18738                  |
| repeat_L1region | /note="L1MD2 repeat: matches 6150. .6335 of consensus" 19070. .19105 |
| repeat_L1region | /note="9 copies 4 mer gata 88% conserved" 19363. .19664              |
| repeat_L1region | /note="L2 repeat: matches 2436. .2750 of consensus" 20032. .20053    |
| repeat_L1region | /note="11 copies 2 mer aa 100% conserved" 21272. .21554              |
| repeat_L1region | /note="A1USX repeat: matches 1. .283 of consensus" 22754. .22918     |
| repeat_L1region | /note="3 copies 55 mer 77% conserved" 26601. .26684                  |
| repeat_L1region | /note="L2 repeat: matches 2668. .2748 of consensus" 26865. .26814    |
| repeat_L1region | /note="A1UJb repeat: matches 2. .131 of consensus" 26815. .27110     |
| repeat_L1region | /note="A1USX repeat: matches 1. .303 of consensus" 27111. .27285     |
| repeat_L1region | /note="A1UJb repeat: matches 130. .307 of consensus" 27286. .27578   |
| repeat_L1region | /note="L2 repeat: matches 2359. .2668 of consensus"                  |

|               |                |  |
|---------------|----------------|--|
| repeat_region | .29945. .30257 | /note="AluI repeat: matches 1. .306 of consensus"      |
| repeat_region | .30812. .31078 | /note="AluXs repeat: matches 1. .267 of consensus"     |
| repeat_region | .31887. .32091 | /note="MER58A repeat: matches 1. .222 of consensus"    |
| repeat_region | .32818. .32853 | /note="9 copies 4 mer acaac 100% conserved"            |
| repeat_region | .32856. .32957 | /note="MER97C repeat: matches 185. .289 of consensus"  |
| repeat_region | .32958. .33406 | /note="MLTIC repeat: matches 1. .466 of consensus"     |
| repeat_region | .33407. .33867 | /note="MER97C repeat: matches 289. .726 of consensus"  |
| repeat_region | .33895. .34718 | /note="8 copies 103 mer 61% conserved"                 |
| repeat_region | .33915. .34519 | /note="121 copies 5 mer atata 56% conserved"           |
| repeat_region | .33920. .34103 | /note="92 copies 2 mer at 57% conserved"               |
| repeat_region | .33924. .34553 | /note="21 copies 30 mer 61% conserved"                 |
| repeat_region | .33925. .34730 | /note="13 copies 62 mer 61% conserved"                 |
| repeat_region | .33940. .34709 | /note="14 copies 55 mer 61% conserved"                 |
| repeat_region | .33954. .34213 | /note="13 copies 20 mer 70% conserved"                 |
| repeat_region | .33999. .34658 | /note="25 copies 28 mer 58% conserved"                 |
| repeat_region | .34014. .34650 | /note="13 copies 49 mer 62% conserved"                 |
| repeat_region | .34071. .34642 | /note="11 copies 52 mer 62% conserved"                 |
| repeat_region | .34300. .34709 | /note="205 copies 2 mer aa 56% conserved"              |
| repeat_region | .34390. .34689 | /note="15 copies 20 mer 60% conserved"                 |
| repeat_region | .34488. .34699 | /note="53 copies 4 mer taac 59% conserved"             |
| repeat_region | .34512. .34703 | /note="64 copies 3 mer ata 60% conserved"              |
| repeat_region | .34629. .34718 | /note="3 copies 30 mer 78% conserved"                  |
| repeat_region | .35062. .35647 | /note="L1ME1 repeat: matches 5550. .6144 of consensus" |
| repeat_region | .35694. .35885 | /note="12 repeat: matches 2054. .2250 of consensus"    |
| repeat_region | .35882. .36229 | /note="12 repeat: matches 1446. .1832 of consensus"    |
| repeat_region | .36377. .36484 | /note="MIR repeat: matches 48. .147 of consensus"      |
| repeat_region | .39838. .39877 | /note="20 copies 2 mer tt 82% conserved"               |
| repeat_region | .39841. .39875 | /note="7 copies 5 mer tggtt 85% conserved"             |
| repeat_region | .41466. .41514 | /note="12 repeat: matches 2699. .2747 of consensus"    |
| repeat_region | .41792. .42143 | /note="THEB1 repeat: matches 1. .364 of consensus"     |
| repeat_region | .43471. .43500 | /note="12 repeat: matches 2629. .2750 of consensus"    |
| repeat_region | .44937. .45157 | /note="12 repeat: matches 2259. .2489 of consensus"    |
| repeat_region | .45491. .45781 |  |

|                       |                |  |          |               |
|-----------------------|----------------|--|----------|---------------|
| Query Match           | 2.1%           | Score 75.4   | DB 9     | Length 143409 |
| Best Local Similarity | 47.6%          | Prod No. 1.4e-05   |          |               |
| Matches 313           | Conservative 0 | Mismatches 341   | Indels 3 | Gaps 3        |
| 0Y                    | 461            | ATTGGTGACATATATAACTTTTGTCTATCTGAAGCTTAATGCTGAGCGAGCGGTTTGGT  | 520      |               |
|                       |                |  |          |               |
| Ob 34685              | 461            | ATTATATATATATATATTTTATTTATATATATATATTTTATTTTATATATATATATTTTA | 34626    |               |
|                       |                |  |          |               |











| Query Match  | 2.08:   | Score 71.4:       | DB 5:           | Length 131033: |
|--|---|-------------------|-----------------|----------------|
| Best Local Similarity                                  | 49.1%:  | Pred. No. 0.0001: |                 |                |
| Matches 189:   | Conservative  | 0:                | Mismatches 196: | Indels 0:      |
| Gaps   | 0:  |                   |                 |                |
| repeat_region  | 16553..16748  |                   |                 |                |
| /note="TDR19 repeat: matches 95..294 of consensus"     |   |                   |                 |                |
| repeat_region  | 16743..16816  |                   |                 |                |
| /note="ANGEL repeat: matches 123..195 of consensus"    |   |                   |                 |                |
| repeat_region  | 16893..17036  |                   |                 |                |
| /note="dir466 repeat: matches 7..148 of consensus"     |   |                   |                 |                |
| repeat_region  | 17182..17355  |                   |                 |                |
| /note="Df000365 repeat: matches 1..182 of consensus"   |   |                   |                 |                |
| repeat_region  | 17508..17812  |                   |                 |                |
| /note="Df000183 repeat: matches 1..315 of consensus"   |   |                   |                 |                |
| misc_feature   | complement(17586..17719)  |                   |                 |                |
| /note="match: STS: Em:545723"                          |   |                   |                 |                |
| repeat_region  | 17721..17869  |                   |                 |                |
| /note="Df000341 repeat: matches 613..763 of consensus" |   |                   |                 |                |
| repeat_region  | 18365..18708  |                   |                 |                |
| Query Match  | 2.08:   | Score 71.4:       | DB 5:           | Length 131033: |
| Best Local Similarity                                  | 49.1%:  | Pred. No. 0.0001: |                 |                |
| Matches 189:   | Conservative  | 0:                | Mismatches 196: | Indels 0:      |
| Gaps   | 0:  |                   |                 |                |
| 612  | ACATAATATATGATGCTATTTATTAAGTCTATGCTATGATCAACTTTCACAGCTCTATACA   | 671               |                 |                |
| 92176  | ACCTTAGATTTATAGCAATAGCTTTTTTTTTTTTGAAGATTTATCAATTTCTATATTTATTTA | 92235             |                 |                |
| 672  | TGTATATATATTTGCGAGTTAGACCTTAATTCAGGTAATTTGTCATTTAATTTATTTATCTG  | 731               |                 |                |
| 92236  | TTTATATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT   | 92295             |                 |                |
| 732  | AATATATATGATGATGCTTTGCTGTTATTTTATGTTTGGTTTCATTTTAAATGACGG       | 791               |                 |                |
| 92296  | ATTTTGTGTAATTTTTTTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT  | 92355             |                 |                |
| 792  | TGAGCTTGTCGATTCATATTTTATTTATGATGACACATCTTTGATGAGATTTATAGAT      | 851               |                 |                |
| 92356  | ATTTATTTTGTGTTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT  | 92415             |                 |                |
| 852  | TGTTAATGCGATGAGGGTTTGCCTGATTTTATTTATTTATTTATTTATTTATTTATTTATTT  | 911               |                 |                |
| 92416  | TAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAAT  | 92475             |                 |                |
| 912  | ATATGTTATTTTGTGCTTTTATATGCTTCTTTTAAAGAGTATGATGACCTTAAGCTC       | 971               |                 |                |
| 92476  | TAAATTTACTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA  | 92535             |                 |                |
| 972  | GCGTAATATGCGGTAATTCGCAT   | 996               |                 |                |
| 92536  | GTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA    | 92560             |                 |                |

RESULT 14

AE009363/c

LOCUS

DEFINITION

Agrobacterium tumefaciens str. C58 linear chromosome, section 133 of 187 of the complete sequence.

ACCESSION

AE009363 AE008689

VERSION

AE009363.1 GI:17742806

KEYWORDS

SOURCE

ORGANISM

Agrobacterium tumefaciens str. C58 (U. Washington).

Agrobacterium tumefaciens str. C58 (U. Washington)

Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.

REFERENCE

1 (bases 1 to 13568)

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida, J.T., N.F., Zhou, Y., Boyce, S.R., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McLellan, E., Palmieri, A., Raymond, C., Rouse, G., Senebittachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nestor, E.W.

The genome of the natural genetic engineer Agrobacterium



```
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL45140.1"
/db_xref="GI:17742813"
/translation="MOFAHVHMLIHVRGGFEMRSLYAVFVPLVATLSGSPAG
AAERTWIFSGALKEGLADPVSDPEMRLLYVAKSAIYAGVADLTSGSDWCGA
GAVAPHELIDRLTYTLTGDMPAKLEDEQATLYREALKVSFPEQKSN"
8744..9244
/gene="Atu4347"
8744..9244
/gene="Atu4347"
8744..9244
/gene="Atu4347"
/note="putative: ORF located using Glimmer"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL45141.1"
/db_xref="GI:17742814"
/translation="MRVNFDTLYSNYPSSDSPHPVYLSQRDLFTEIGWESFIGNPNH
MTCAIRSIAFVKSIGINIVPSHSRIQKPYAGKGIENVMMRLATIMKRTSYLGEEDPY
TPATARNIGIGARNVAVANNIPGTYGGCHIDLVRGSEATQCASACIYNSETIHWPL
QASRGS"
9239..11709
/gene="vgrg"
9239..11709
/note="vgrg"
9239..11709
/note="vgrg"
/note="identified by sequence similarity; putative: ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="vgrg protein"
```

Query Match 2.0% Score 70.8; DB 1; Length 13568;  
Best Local Similarity 52.9%; Pred. No. 0.00014;  
Matches 175; Conservative 0; Mismatches 153; Indels 3; Gaps 1;

```
QY 3258 AGCGCTAGTAAAGACGCTGTCACATCATGCTGATGCTGAACGAATGGCGCCGTAA 3317
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3657 AGCGGTTGGCAACACCCGCTTACATTAATACGAAGTCGAACCGAAGCGCGCATGG 3598
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3318 AGACGAAGAGCTGCCGTTGTTGGCGTCATTGCGACATTTTTCAGGACACMAACGAG 3377
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3597 TGGTCAAGGAAGCTTCTTCTAGTAGTGGCGTGCGTGATTTCTCCGGCAATCCACCC 3538
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3378 AATCAGAAAAAGTTGATTAGAACGACGAGTTCACGGGTATCGATTAAAGACACTTGG 3437
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3537 AGCCGCTCAAGCGC---TTCCGTTGAACGCCAAATTCGTTGAGATCGACCGCACAATTTCG 3481
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3438 ATACACTGATGGGCAAAATTCACCCGCGTTCGTACAAAGTTGATTAACAAGCTTGTA 3497
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3480 ACGAAGTCATGCCCGCCGATGACGCCCGGCTCAACATTTCCGTTGCCAACACGCTGCA 3421
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3498 ATGATGATAGCAGTTTGAAGTGAAGTGAAGCTCCGTTGATGAAGAAATTTCCACCCAG 3557
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3420 AAGACGCGACGACATGACGTCACCTGAAGTTGCAAAAGCANTGACGATTTCCAAACCCG 3361
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3558 AGAAGTTAGTTGATTAATGAGCCGCTTAA 3588
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3360 GCGCCGTTGTGACGAGGATTCGCGCGCTCA 3330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 15  
AE008251  
LOCUS AE008251 1961 bp DNA linear BCT 18-DEC-2001  
DEFINITION Agrobacterium tumefaciens str. C58 linear chromosome, section 55 of  
187 of the complete sequence.  
ACCESSION AE008251 AE007870  
VERSION AE008251.1 GI:15158900  
KEYWORDS  
SOURCE Agrobacterium tumefaciens str. C58 (Cereon).  
ORGANISM Agrobacterium tumefaciens str. C58 (Cereon)  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

```
REFERENCE  
AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.  
TITLE Complete Genome Sequence of Agrobacterium tumefaciens C58  
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall  
Disease in Plants  
JOURNAL Unpublished  
2 (bases 1 to 1961)  
AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.  
TITLE Direct Submission  
JOURNAL Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney  
Street, Cambridge, MA 02139, USA  
COMMENT Approximately 800 bp of telomeric sequence missing from the left  
end of the chromosome and 200 bp missing from the right end.  
FEATURES  
source location/Qualifiers  
1..1961  
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"  
/strain="C58"  
/db_xref="taxon:181661"  
183..1238  
/gene="AGR_L_1042"  
183..1238  
/gene="AGR_L_1042"  
/note="hypothetical protein PA2360 (Imported) -  
pseudomonas aeruginosa (strain PA01)"  
/codon_start=1  
/transl_table=11  
/product="AGR_L_1042p"  
/protein_id="AAK89097.1"  
/db_xref="GI:15158901"  
/translation="MDTOHVKRDIEFLNCGDNIRNDSRTREIYRINDERQARAE  
RAASPDNLKISSWDSVSNGLDIYESKQVELAWLAESLIRGFHGRREYEL  
CGDIFNHDMSLRISDNDDEKFAFPAGLNGISGGLVLOPLRSLIIPGKFEHS  
LMDPOLARPNSKRRELYRTASAGVAAASHLAAYNTCISLSDATITAYLSECCG  
AAPSSNTRNTLIEAAAIRTLGGDDPAPVEQTPALAGTDESGQSAARTSPASPE  
GISRDEAFETLSVARYFRTEPHSPISLSETLVRRGRMDFSELLAELPETQARN  
AVLTAAIGIKPGGDNNGK"  
1277..1813  
/gene="AGR_L_1045"  
1277..1813  
/gene="AGR_L_1045"  
/note="conserved hypothetical protein PA2365 (Imported) -  
pseudomonas aeruginosa (strain PA01)"  
/codon_start=1  
/transl_table=11  
/product="AGR_L_1045p"  
/protein_id="AAK89098.1"  
/db_xref="GI:15158902"  
/translation="MPRGKEVEMASVHEKLEVRKPRVHIKEVEYEGAMVKELEP  
VVGVLDFSGNPTOLKPRGGERKPVQIDRDNDDEVMRRMTEGLANTSVANTLOKDTM  
QYNLFESMDPEPGAVYQVYALKAALLDARNELDLISKADRSDDLRLMEDIQNK  
TDLATVSELKENGASE"  
1907..3388  
/gene="AGR_L_1047"  
1907..3388  
/gene="AGR_L_1047"  
/note="(Af037441) putative 54.5 kda protein"  
/codon_start=1  
/transl_table=11  
/product="AGR_L_1047p"  
/protein_id="AAK89099.1"  
/db_xref="GI:15158903"  
/translation="MSASLSLKSEAOATTAENOGLSKVSATROTEPPDRONLRLTL  
TDQALGTVKTDRLNLTJLNLHIAELDRTISQDLAIAQAPFAKLEGTWROLNLYLK  
NSETSVNLKIRVMNAGKRELADELKAVEPDOSRLFKAIYEDFETPGEGPLCATIGD  
YEFNGSPDVOLQGVSSIAAFAFPITSAAPIMHGFEDYRDLRPPDLKIDFETVE  
YAKWRSFRESDDSRVTLALDPLVLAIRMYGCPNTVDDFAVDETGAVNGDLQHDREYC  
WMAAYVMGTGKLTLEAFKSGMCTAIRGENGGRVNLPMHAYSSDDGLDKCPREVG  
ITDRDAELKLGRLPLCHYKKNTOYAVFPGNOTAIPKLYOKPEATANAAYASARPIYI  
MATSFRFHTLKWGRDKTIGFMEASDCVWILRWIANTVNADEAGEBSRAATYPLRDA  
KVTVGEVPGKPGAYNAVAMRPMLOMBELTJSLRVAVARIPSKN"  
3401..4837  
/gene="AGR_L_1048"
```



Mon Mar 3 10:44:09 2003

us-09-915-706a-1.rge

Page 17

Search completed: March 2, 2003, 02:21:44  
Job time : 15022 secs

---





PT useful for detecting cytosine methylation state of genes associated  
PT with pharmacogenomics and for therapy of diseases e.g. cancer  
PS Claim 1; SEQ ID No 18; 24pp; English.

XX  
CC The invention relates to a nucleic acid comprising a sequence at  
CC least 18 bases in length of a segment of the chemically pretreated DNA  
CC of genes associated with pharmacogenomics according to one of the  
CC sequences of the genes ALDH6 (NM\_000693), CYP11A (NM\_000781), CYP11B1  
CC (NM\_000497), CYP3A3 (NM\_000776 and NM\_017460), DPYD (NM\_000110), EPHX2  
CC (NM\_001979), OCIN (NM\_002538), TXNRD1 (NM\_003330), UGT8 (NM\_003360),  
CC MRP (NM\_004996, NM\_019900, NM\_019901, NM\_019902, NM\_019862, NM\_019898,  
CC NM\_019899) and their complementary sequences, or a sequence (SI) chosen  
CC from 87 sequences and their complements. The chemical pretreatment  
CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)  
CC into uracils. Also included are an oligomer (II) in particular an  
CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in  
CC each case at least one base sequence having a length of 9 nucleotides  
CC which hybridises to or is identical to a chemically pretreated DNA of  
CC genes associated with pharmacogenomics and their complements, arranged in  
CC an array for analysing diseases associated with the methylation state  
CC (CPG) and/or detecting SNPs (single nucleotide polymorphisms)  
CC of the 87 sequences. The oligomers may also be used as PCR primers.  
CC The set of 87 nucleic acids and their complements is useful for diagnosis  
CC and therapy of solid tumours and cancer. The present sequence  
CC represents one the 87 DNA sequences or its complement.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcl\_sequences.

XX  
SQ Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 other;

Query Match 2.0%; Score 72.8; DB 24; Length 11422;

Best Local Similarity 47.7%; Pred. No. 4.1e-07;  
Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;

QY 605 AATTATACATAAATATGATGCTATTATTAATCTGCTGATATGATACACTTCAGCT 664  
DB 9660 AATTAAATTTTAAATGATGCTATGCTATGATATATAGTAATATAGTGTAGT 9719  
QY 665 CTATACATGTAATATTTGAGTACCTTAATTCAGGTATTTGCTATTATTA 724  
DB 9720 TAAATGGTTATATATATTTTGTATTTTATATATATATATTTTATTTAA 9779  
QY 725 TTATCTGAATATATGTAATGATGCTTGTGCTATTTATTTATGTTTCAATTTTA 784  
DB 9780 TATTTTGAATATTTGCTGATATATATATATATATGCTATATATGTTTATTTT 9839  
QY 785 ATGACGGTGAAGCTTGTGCATTCATATTTTATGATGACCAACATCTTGATGAATTT 844  
DB 9840 GTTATTTATATGATGATTAATATATATTTAGTAAATATATTTTGAAGTATTTT 9899  
QY 845 AAGATATTTGTAATGATGAGGGTTGGCGTATTT-TTATATTTAATCATATAA 903  
DB 9900 TATTTATTTGAATATATGATTTTATTTAAATTTATGATTTATTTTAAATGATTA 9959  
QY 904 TCAACATATATATGTTATTTGCTTTTATATAGTGTCTTTTAAAGAGATGATGACC 963  
DB 9960 TTATTTATTTTATTTTATTTTATTTTGAATTTTAAAGACATTTAAGTAAATTTGTTT 10019  
QY 964 TAAAGTGGCTAAATATGCGTAAATTTGCAATGCTATATATTCACCTCAAGATACACT 1023  
DB 10020 TAAATTTATATGATATCGAGGTAGTTTTCATTTTATTTTATATATATTAATAGGTA 10079  
QY 1024 ATTGGCAATTTGCAATATGCTCGTATGCAACAAATTTGATGATGCTGTTTGG 1083  
DB 10080 TTTTGTATATGATATTTGATGATATTTTAAATTTTATTTTATTTTATTTTGA 10139  
QY 1084 CTGCAAAATATTAATTTTCTGTTGAATA 1115  
DB 10140 TTGCAATATATTAATATGATATGATTAAGTA 10171

RESULT 2  
ABL32219  
ID ABL32219 standard; DNA; 11422 BP.

XX  
AC ABL32219;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 192.

XX  
KW Human: immune system disease; cytosine methylation; antiasthmatic;  
KW antidiabetic; antianemic; cytosolic; noctropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antineoplastic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

XX  
OS Homo sapiens.

XX  
PN WO200200928-A2.

XX  
PD 03-JAN-2002.

XX  
PF 02-JUL-2001; 2001WO-EP07537.

XX  
PR 30-JUN-2000; 2000DE-1032529.

XX  
PR 01-SEP-2000; 2000DE-1043826.

XX  
PI (EPIG-) EPIGENOMICS AG.

XX  
PA Olek A, Piepenbrock C, Berlin K;

XX  
PI WPI; 2002-130909/17.

XX  
PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -

XX  
PS Claim 1; SEQ ID NO 192; 32pp + Sequence Listing; German.

XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX  
SQ Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 other;

Query Match 2.0%; Score 72.8; DB 24; Length 11422;

Best Local Similarity 47.7%; Pred. No. 4.1e-07;  
Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;

QY 605 AATTATACATAAATATGATGCTATTATTAATCTGCTGATATGATACACTTCAGCT 664  
DB 9660 AATTAAATTTTAAATGATGCTATGCTATGATATATAGTAATATAGTGTAGT 9719  
QY 665 CTATACATGTAATATTTGAGTACCTTAATTCAGGTATTTGCTATTATTA 724  
DB 9720 TAAATGGTTATATATTTTGTATTTTATATATATATATTTTATTTTAA 9779  
QY 725 TTATCTGAATATATGTAATGATGCTTGTGCTATTTTATGTTTCAATTTTA 784  
DB 9780 TATTTTGAATATTTGCTGATATATATATATATATGCTATATATGTTTATTTT 9839  
QY 785 ATGACGGTGAAGCTTGTGCATTCATATTTTATGATGACCAACATCTTGATGAATTT 844  
DB 9840 GTTATTTATATGATGATTAATATATTTTAGTAAATATATTTTGAAGTATTTT 9899



XX 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2001-602752/68.  
 XX  
 XX Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for  
 PT analysing diseases associated with cytosine methylation state e.g.  
 cancer  
 PS  
 PS Claim 1; SEQ ID No 158; 27pp; English.  
 CC The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 CC (5s) and sequences complementary to (5s). The nucleic acid may be a  
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with CpG dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC and/or therapy of existing diseases or the predisposition to specific  
 CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and  
 CC oncogenes. Sequences with even numbered Seq ID numbers are the  
 CC complementary sequence of the corresponding odd numbered sequence (e.g.  
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
 CC is missing).  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 other:  
 SO  
 Query Match 1.9%; Score 69.4; DB 22; Length 8093;  
 Best Local Similarity 47.5%; Pred. No. 2.4e-06;  
 Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

OY 827 ATCTTGATGAGTATTTAAGATATGTTAATGCATGAGGGCTTGGCTATTTTAT 886  
 DB 6585 GGATTTAATTTATATATATATATTTTAAAGTATATATATTTT-AAAAATTTTGT 6642  
 OY 887 ATTTAATCATATATAAATCAACAAATATATGTTATGCTTTTATAGTCTTTT 946  
 DB 6643 TTTAAGCTTTTGTATATATATGAAATATTTTATTTGCTTTATATAGTAAATTT 6702  
 OY 947 AAGAGGTAGATGACCTAAAGCTGCGCTAATATATGCGTAAATTTGCATTTATAT 1006  
 DB 6703 AAGCAATTAGCTATATGATTAAG-----AATTTAGGTGAATGATGTTGGGTAT 6754  
 OY 1007 CACCTCAAGATACATATTTGGCAATTTGCAATATGCTCGATGCAACAATTT 1066  
 DB 6755 AAAATTAATGCTATATAGAGTTTAAAGTTGCTTATGCTTTTATATAGAGT 6814  
 OY 1067 AGTAGATGTTTGTGCTGCAAAAATTAATTTTCTGCT 1109  
 DB 6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTAGGGGT 6857  
 RESULT 5  
 ABL92237  
 ID ABL92237 standard; DNA; 8093 BP.  
 XX  
 AC ABL92237;  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Chemically treated DNA repair gene fragment complementary to#23.  
 XX  
 XX DNA repair: cytosine methylation; PMS2L1; PMS2L12; PMS2L3;  
 KW PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4;  
 KW DITTL1; FANCB; XRC8; ataxia telangiectasia; aging; Bloom's syndrome;  
 KW Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;  
 KW immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;  
 cancer; ds.  
 KW  
 OS Unidentified.  
 XX  
 PN WO200181622-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 06-APR-2001; 2001WO-EP03972.  
 XX  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2002-03446/04.  
 XX  
 PT New nucleic acid derived from genes associated with DNA repair, useful  
 PT for diagnosis, e.g. of ataxia telangiectasia, by determination of  
 PT cytosine methylation  
 XX  
 PS Claim 1; SEQ ID NO 46; 25pp + sequence listing; English.  
 CC The invention relates to nucleic acids containing a sequence of at least  
 CC 18 nucleotides of chemically treated DNA of genes associated with DNA  
 CC repair, and their complements. The invention also relates to nucleic  
 CC acids comprising at least 18 base pairs of the chemically pretreated DNA  
 CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,  
 CC PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,  
 CC RFC4, DITTL1, FANCB, or XRC8. Nucleic acids of the invention and related  
 CC oligomers, are useful for diagnosis of diseases associated with gene  
 CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,  
 CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,



```
QY 587 TATTGGATGATTCGAAATTCACATATAATGCTGCTATTTACTTGGCTGATT 646
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6345 GTATTAGATTTTGTAGATGTTAGAAATGATTTATTTAAATTTGGAATGCGTATA 6404
QY 647 GGTATCAACTTTCAGCTCTATACATGTAATATATTTTCGAGTTCGACCTTAATTCACAGCT 706
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6405 TATTTTATGCTAAGATAGATATTAAGTAAATATTTAAAGATGTTTATATAGAT 6464
QY 707 AATTTGCTATTTATTTATTTATGTAATTAATGTAATGCTGTTTGTGCTATTTT 766
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6465 TGTAGTAATTTTGTATTTTATTTAAGATTAATTTGTTTGTATTAATAGTAATTTTAA 6524
QY 767 ATGTTTGTTCATTTTAAATGACGCGTTCGATTCATTAATTTTATGATGACAC 826
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6525 ATTTTGTGTTATTTGAAGAAAGTAAATTTTAAAGTTTATATGTAATTTATTAATA 6584
QY 827 ATCTTGTATGAGATTTTAAAGATTTGTTAATGATGACGCGTTTGGCTGATTTTAT 886
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6585 GCATTTAATTTATTTATATAGATTTTATTAAGTATTTATTAATTTT -AAAAATTTTGT 6642
QY 887 ATTTAATCATATTAATTAATCAATATATGTTATTTTGTCTTTTATAGTTCCTTT 946
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6643 TTAAAGTTTGTTTTATTTATGAAATATTTTAATTTGTGTTATTAATTTAGTAATTT 6702
QY 947 AAAGAGGTAGATGACCTAAAGTCCGCTAATATGCGTAATTTGCCATTCATTAAT 1006
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6703 AAGGAATTTAGTTATGATTTAG-----AATTTAGGTGAATTTGATTTGGGTATTT 6754
QY 1007 CACCTCAAGATACACTATTTGGCAATTTGACAAATATGTCCTTCGATGCAACAATATT 1066
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6755 AAAATTAATGTTATTAAGGTTTAAAGTTTAAAGTTTGTGTTTAAATTTAGAGGT 6814
QY 1067 AGTAGATGTTTGTGTTGCTGCAAAATTAATTTTCTGCTT 1109
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTTAGGGGT 6857

RESULT 7
ABL49332
ID ABL49332 standard; DNA: 8093 BP.
XX
AC ABL49332;
XX
DT 01-MAY-2002 (first entry)
XX
DE Human polynucleotide associated with DNA replication SEQ ID NO 32.
XX
KW Human: cytosolic; neuroprotective; nootropic; immunostimulant;
KW gene therapy; gene regulation; DNA replication; CENPB; DNA2L; ATR; CHD1L;
KW ERCC3; SNRPB1; RAD50; Lig2; cytosine methylation; Ataxia telangiectasia;
KW ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200177377-A2.
XX
PD 18-OCT-2001.
XX
PE 06-APR-2001; 2001WO-EP03971.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017471/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
```

```
PT with DNA replication, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. Ataxia telangiectasia
PS Claim 1; SEQ ID NO 32; 23pp + Sequence Listing; English.
XX
CC The invention relates to nucleic acid sequences comprising at least 18
CC bases of a chemically pretreated gene associated with gene regulation,
CC selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences
CC associated with DNA replication, CENPB, DNA2L, ATR, CHD1L, ERCC3, SNRPB1,
CC RAD50 and Lig2. The chemical pretreatment converts cytosine bases
CC unmethylated at the 5-position to uracil or another base with
CC hybridisation behaviour dissimilar to cytosine, to enable analysis of
CC cytosine methylations. The DNA sequences and method are useful in the
CC diagnosis of diseases (or predisposition to diseases) associated with DNA
CC replication and in therapy of such diseases, by enabling analysis of the
CC cytosine methylation patterns of such genes. They are especially useful
CC in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's
CC syndrome, solid tumours and cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification and was supposed to be available directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. However, the sequence data did
CC not correspond to that referred to in the specification. The present data
CC is taken from EPO data for the patent.
XX
SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 other:
XX
Query Match 1.9%; Score 69.4; DB 24; Length 8093;
Best Local Similarity 47.5%; Pred. No. 2.4e-06;
Matches 277; Conservatively 0; Mismatches 296; Indels 10; Gaps 2;
QY 527 ATTCGTTACGTCATATTCATTAATTCATTAATTAATTAACAGTCTAAAGCGTGT 586
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6285 ATGCTGCTTTTATTTTAAATTTTGTTTTAAAGTAATATTTGTTTAAAGATATA 6344
QY 587 TATTGGATGATTTCCGAATTCACATATATGATGCTATTTACTTGGCTGATT 646
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6345 GTATTAGATTTTGTAGATGTTAGAAATGGAATTTATTTTAAATTTGGAATTCGCTATA 6404
QY 647 GGTATCAACTTTCAGCTCTATACATGTAATATATTTTGAAGTTAGACCTTAATTCAGAGT 706
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6405 TATTTTATGCTAAGATAGATATATAGAGAAATTTTAAAGATGTTTATTTAATTAAGAT 6464
QY 707 AATTTGCTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 766
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6465 TGTAGTAATTTTGTATTTTATTAAGATTAATTTGTTTGTCTTAAATTAATTTTAA 6524
QY 767 ATGTTTGTTCATTTTAAATGACGCGTTCGATTCATTAATTTTATGATGACAC 826
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6525 ATTTTGTGTTATTTGAAGAAAGTAAATTTTAAAGTTTATATGTAATTTATTAATA 6584
QY 827 ATCTTGTATGAGATTTTAAAGATTTGTTAATGATGAGGCGTTTGGCTGATTTTAT 886
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6585 GCATTTAATTTATTTATATAGATTTTATTAAGTATTTATTAATTTT -AAAAATTTTGT 6642
QY 887 ATTTAATCATATTAATTAATCAATATATGTTATTTGTCTTTTATAGTTCCTTT 946
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6643 TTAAAGTTTGTTTTAAATTTTGAATATTTTATTTGTGTTTATTAATTTAGTAATTT 6702
QY 947 AAAGAGGTAGATGACCTAAAGTCCGCTAATATGCGTAATTTGCCATTCATTAATTT 1006
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6703 AAGGAATTTAGTTATGATTTAG-----AATTTAGGTGAATTTGATTTGGGTATTT 6754
QY 1007 CACCTCAAGATACACTATTTGGCAATTTGACAAATATGTCCTTCGATGCAACAATATT 1066
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6755 AAAATTAATGTTATTAAGGTTTAAAGTTTAAAGTTTGTGTTTAAATTTAGAGGT 6814
QY 1067 AGTAGATGTTTGTGTTGCTGCAAAATTAATTTTCTGCTT 1109
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTTAGGGGT 6857

RESULT 8
ABL70350
```





|    |       |  |       |
|----|-------|--|-------|
| QY | 867   | GGTTGGCTGATATTTTATATATTAATCATATAAATCAACAATATATGTTATTTGTC       | 926   |
|    |       |  |       |
| Db | 56941 | GATTTATATTTATATATATATATATATAAATTTAAATGTTAATAATTTTTTTGTTTATATGT | 57000 |
| QY | 927   | TCTTTTATAGTGTTCTTTTAAAGAGTAGATGACCTAAAGGTGCGCTAAATATGGCGT      | 986   |
|    |       |  |       |
| Db | 57001 | TTATATAGTAGGAAGCTTATACGTATATGTTGATGTGATTTTAAATTTTATATATTTAT    | 57060 |
| QY | 987   | AAATTCGCAATGCTATTAATTCACCTCAAGATPACAGTATGGCAATTTGACAAATATGTC   | 1046  |
|    |       |  |       |
| Db | 57061 | AAATTTTAGCTATGTTTATATATGTTGGATTAATTTTAAATATTTTGTAGTATATTT      | 57120 |
| QY | 1047  | ACTTCGTA----TGAAACAATATAGTAGAGTGCTTTTGCTCGCAAAATPAAATTTT       | 1102  |
|    |       |  |       |
| Db | 57121 | ATTTTATGTTTATTAATAATGACGATTTATTTTTTTTTTTTTTTTTTATATTTTATATTT   | 57180 |
| QY | 1103  | TCTGGTGAATAATCACTCAAGCGCTACGCTTTCCTTATCTPAAATATACAGAAATAG      | 1162  |
|    |       |  |       |
| Db | 57181 | TTTTTTTATTTTGGTTATGATTTTATTTTATTTTGGGAAATATAGAAAGTAAATA        | 57240 |
| QY | 1163  | CGATTGAAGTTAATTTGACACTTAACCAATATGTCACCTAACAGACAGGAACCTATGCC    | 1222  |
|    |       |  |       |
| Db | 57241 | AAAT--GAGTTGTTATTTTATTAATAAATTTATATAGTTATTTTATTTTATTTATATG     | 57298 |
| QY | 1223  | TTTGTCAAGCATCAATATGAGCAACTCTTAACCTCTGAGTGTGATGTGATCTGTGG       | 1282  |
|    |       |  |       |
| Db | 57299 | TTTATTTTTTTTTTAGATTTTGAAGATTTATTTTATTTTGAAGCATTAATTTGGATTTATA  | 57358 |
| QY | 1283  | CGTTTATCTTAACCT 1297   |       |
|    |       |  |       |
| Db | 57359 | TGTTTATTTTAAATTT 57373   |       |
|    |       |  |       |

|           |   |
|-----------|---|
| RESULT 11 |   |
| ID        | AAS45516  |
|           | AAS45516 standard; DNA; 6306 BP.  |
| XX        |   |
| AC        | AAS45516;   |
| XX        |   |
| DT        | 18-DEC-2001 (first entry)   |
| XX        |   |
| DE        | Chemically pretreated genomic DNA associated with cell cycle #11.         |
| XX        |   |
| KW        | Cell cycle; human; CGP dinucleotide; cytosine methylation; HIV; aging;    |
| KW        | human immunodeficiency virus; neurodegenerative disorder; solid tumour;   |
| KW        | graft-versus-host disease; glomerular disease; Lewy body disease; cancer; |
| KW        | arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;    |
| KW        | immunosuppressive; antitumour; cyostatic; antiarteriosclerotic; ds;       |
| KW        | PCR primer.   |
| XX        |   |
| OS        | Homo sapiens.   |
| XX        |   |
| PN        | WO200168911-A2.   |
| XX        |   |
| PD        | 20-SEP-2001.  |
| XX        |   |
| PF        | 15-MAR-2001; 2001WO-EP02945.  |
| XX        |   |
| PR        | 15-MAR-2000; 2000DE-1013847.  |
| PR        | 06-APR-2000; 2000DE-1019058.  |
| PR        | 07-APR-2000; 2000DE-1019173.  |
| PR        | 30-JUN-2000; 2000DE-1032529.  |
| PR        | 01-SEP-2000; 2000DE-1043826.  |
| XX        |   |
| PA        | (EPig-) EPIGENOMICS AG.   |
| XX        |   |
| PI        | Olek A, Piepenbrock C, Berlin K;  |
| XX        |   |
| DR        | WPI; 2001-602751/68.  |
| XX        |   |
| PT        | Designing primers and probes for analysing diseases associated with       |
| CT        | cytosine methylation state e.g. arthritis, cancer, aging.                 |

PT arterioleclerosis comprising fragments of chemically modified genes  
PT associated with cell cycle -  
XX  
XX  
XX Claim 1; SEQ ID No 221; 28pp; English.  
XX  
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
CC molecules associated with the cell cycle and specific PCR primers of the  
CC invention. The sequences are useful for detecting the methylation state  
CC of all CPG dinucleotides in a sequence and therefore for analysing  
CC associated diseases. By analysing cytosine methylations in the pretreated  
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
CC of existing diseases or the predisposition to specific diseases can be  
CC ascertained. The parameters may be compared to another set of genetic  
CC and/or epigenetic parameters, the differences serving as basis for  
CC diagnosis and/or prognosis events which are disadvantageous to patients.  
CC The sequences of the invention are useful for the diagnosis and therapy  
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
CC aging, glomerular disease, Lewy body disease, arthritis,  
CC arteriosclerosis, solid tumours and cancers.  
XX  
XX Sequence 6306 BP; 1822 A; 190 C; 1449 G; 2845 T; 0 other;  
XX

|                           |        |                    |           |              |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match               | 1.9%   | Score 68;          | DB 22;    | Length 6306; |
| Best Local Similarity     | 49.8%; | Pred. No. 4.8e-06; |           |              |
| Matches 201; Conservative | 0;     | Mismatches 200;    | Indels 3; | Gaps 1       |

|    |      |  |              |     |
|----|------|--|--------------|-----|
| QY | 586  | TTATTGATGAATAATTCGAATAATTCACATAATAATGATGCTAATTAATTA              | ACTTGCTGTAAT | 643 |
| Db | 1745 | TTATTTTATTTATTTTCGTAATTTTATTTATTCGTTAATTTTATTTATTTTATTTAT        | 1804         |     |
| OY | 646  | TGGATACACTTCATCCCTCATACANGTAAATTAATTCGATGATGACCTTAATTCAGG        | 705          |     |
| Db | 1805 | TTTTTTTATTTTATTTTATTTTATTTATTCGTATATTTTACGGCGTATTTTAAATTTTAT     | 1866         |     |
| QY | 706  | TAAATTCCTATTTTAAATTAATTAATCTGAATAATATGTAAATCGATTCGTTGTGTAATTTT   | 765          |     |
| Db | 1865 | TTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT    | 1924         |     |
| OY | 766  | TAATCTTTGTTTCATTTTAAATGACGGTGACCTTGCCATTCATATTTTATATGATGACAA     | 825          |     |
| Db | 1925 | TTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT   | 1984         |     |
| OY | 826  | CATCTTGGATGAGCATTTAAGATATTTGTTAAATGCATGAGGGGTTTGCCTGATATTTT - -  | 883          |     |
| Db | 1985 | TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT | 2044         |     |
| OY | 884  | -TATATTAATTCATTAATAAATCAACAATAATATGTAATTTTGTCTTTTATTAAGTTC       | 942          |     |
| Db | 2045 | ATTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT   | 2104         |     |
| OY | 943  | TTTTTAAGAGGTAGATGACCTTAAGAGTGCCCTAAATATAGCGCT                    | 986          |     |
| Db | 2105 | TTTTTAAGGAAGTAAGTATTTTTCATTCGGAATAATTAAGAGCT                     | 2148         |     |

RESULT 12  
ABK28458  
XX ABK28458 standard; DNA: 6306 BP.  
AC  
XX ABK28458;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE DNA transcription associated complementary genomic DNA #166.  
XX  
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
KW immunological disorder; Werner syndrome; developmental disorder;  
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;  
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;  
KW megaloblastic syndrome; myocardial infarction; hypertension; arthritis

KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;  
KM polyglutamine disorder; solid tumour.  
XX Unidentified.  
OS  
XX WO200192565-A2.  
XX  
XX 06-DEC-2001.  
XX  
XX 06-APR-2001; 2001WO-EP03973.  
XX  
XX 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-090046/12.  
XX  
XX  
XX New nucleic acids or oligomers, useful for diagnosing or treating  
PT diseases associated with DNA transcription, e.g. immunological  
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
PT tumours or cancer -  
XX  
XX  
XX Claim 1; SEQ ID NO 332; 32pp; English.  
XX  
XX The invention relates to a nucleic acid, which comprises a segment of the  
XX chemically pretreated DNA of genes associated with DNA transcription from  
XX one of 346 sequences, and an oligomer, in particular an oligonucleotide  
XX or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
XX to the chemically pretreated DNA of genes associated with DNA  
XX transcription. The set of oligomer probes are useful for detecting the  
XX cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
XX in a chemically pretreated genomic DNA. The nucleic acids are useful for  
XX diagnosing or treating diseases associated with DNA transcription  
XX (particularly with the methylation status), e.g. adenosine deaminase  
XX deficiency, viral infection, retroviral infection, Sezary syndrome,  
XX hematological disorders, immunological disorders, Werner syndrome,  
XX tubercolusis, developmental disorders, psoriasis, Rieger's syndrome,  
XX neurological disorders, neurodegenerative disorders, Waardenburg  
XX syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
XX infarction, hypertension, angiogenesis, erythropoiesis, congenital heart  
XX disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
XX or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
XX associated genomic DNA molecules of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format directly from the  
XX European Patent Office.  
XX  
XX Sequence 6306 BP; 1822 A; 190 C; 1449 G; 2845 T; 0 other:  
SQ  
Query Match 1.9%; Score 68; DB 24; Length 6306;  
Best Local Similarity 49.8%; Pred. No. 4.8e-06;  
Matches 201; Conservative 0; Mismatches 200; Indels 3; Gaps 1;  
QY 586 TTATTCGATGATATTCGAATATTCACATATATTCATGCTTATTCCTGCTAT 645  
DB 1745 TTATTTTATTTATTCGTTATTTTATTCGTTATTTTATTTTATTTTATTTTAT 1804  
QY 646 TCGATACACTTCGATGCTGATACATATATTCGATGATGACCTTAATTCAGG 705  
DB 1805 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1864  
QY 706 TAATTTCTATTTATTTATTTATTCGAATATATGATATGATGCTTTGGTTATTTT 765  
DB 1865 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1924  
QY 766 TATGTTGTTTCATTTTATTAATGACGGAGCTGTGCATTCATATTTTATGATGACA 825  
DB 1925 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1984

QY 826 CATCTTGTGATGACATATTTAAGATATTTGTAATGACATGAGGGCTTGCTGATTTT -- 883  
DB 1985 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2044  
QY 884 -TATATTAAATCATATTAATAACACATATATGATTTTGGCTTTTATGATGTTTC 942  
DB 2045 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2104  
QY 943 TTTTAAAGAGGTAGGATGACCTTAAGGCTGCTTAATATGCGGT 986  
DB 2105 TTTTAAAGAGGTAGGATGATTTTTCGATCGGAAATTAAGACGT 2148  
RESULT 13  
ID ABL92249 standard; DNA: 13131 BP.  
XX  
XX ABL92249;  
XX  
XX 01-JUL-2002 (first entry)  
XX  
XX  
XX Chemically treated DNA repair gene fragment complementary to#29.  
XX  
XX DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L2; PMS2L3;  
KM PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDP1; TDG; INPPL1; RFC4;  
KM DITFL; FANCB; XRC8; ataxia telangiectasia; aging; Bloom's syndrome;  
KM Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;  
KM immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;  
XX cancer; ds.  
XX  
XX Unidentified.  
XX  
XX WO200181622-A2.  
XX  
XX 01-NOV-2001.  
XX  
XX 06-APR-2001; 2001WO-EP03972.  
XX  
XX 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-034446/04.  
XX  
XX  
XX New nucleic acid derived from genes associated with DNA repair, useful  
PT for diagnosis, e.g. of ataxia telangiectasia, by determination of  
PT cytosine methylation -  
XX  
XX  
XX Claim 1; SEQ ID NO 58; 25pp + sequence listing; English.  
XX  
XX The invention relates to nucleic acids containing a sequence of at least  
XX 18 nucleotides of chemically treated DNA of genes associated with DNA  
XX repair, and their complements. The invention also relates to nucleic  
XX acids comprising at least 18 base pairs of the chemically pretreated DNA  
XX of genes associated with DNA repair selected from PMS2L1, PMS2L12,  
XX PMS2L2, PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDP1, TDG, INPPL1,  
XX RFC4, DITFL, FANCB, or XRC8. Nucleic acids of the invention and related  
XX oligomers, are useful for diagnosis of diseases associated with gene  
XX repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,  
XX Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,  
XX immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours  
XX and cancer, particularly by determining status of cytosine methylation  
XX and/or by detecting single-nucleotide polymorphisms. Determination of  
XX individual methylation patterns may allow development of individualised  
XX therapies. The sequences given in records ABL92192-ABL92335 represent  
XX chemically pre-treated DNA fragments from genes associated with DNA  
XX repair, and their complements.





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 16:46:40 : Search time 120 Seconds  
(without alignments)  
9169.635 Million cell updates/sec

Title: US-09-915-706A-1

Perfect score: 3588

Sequence: 1 gtcgacttattgcattgatg.....gatnaattgcgcgcttaa 3588

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 443362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/lna/PCrUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/lna/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 68.6  | 1.9         | 837    | 4  | US-08-998-416-288  |
| 2          | 66    | 1.8         | 19124  | 2  | US-08-487-826B-13  |
| 3          | 65    | 1.8         | 658    | 4  | US-08-998-416-595  |
| 4          | 59.2  | 1.6         | 665    | 2  | US-08-883-795A-36  |
| 5          | 55.6  | 1.5         | 8920   | 2  | US-08-446-855A-1   |
| 6          | 55.6  | 1.5         | 8920   | 4  | US-09-150-741-1    |
| 7          | 53    | 1.5         | 19124  | 2  | US-08-487-826B-13  |
| 8          | 52.8  | 1.5         | 6243   | 2  | US-09-056-075-1    |
| 9          | 49.8  | 1.4         | 636    | 4  | US-08-998-416-186  |
| 10         | 48.4  | 1.3         | 615    | 4  | US-08-998-416-1137 |
| 11         | 48    | 1.3         | 168575 | 4  | US-09-426-290-1    |
| 12         | 46.6  | 1.3         | 5852   | 1  | US-07-867-106-2    |
| 13         | 44.8  | 1.2         | 662    | 4  | US-08-998-416-185  |
| 14         | 44.8  | 1.2         | 665    | 4  | US-08-998-416-937  |
| 15         | 44.8  | 1.2         | 2110   | 4  | US-09-419-459-1    |
| 16         | 44.6  | 1.2         | 4285   | 4  | US-09-410-464-1    |
| 17         | 44    | 1.2         | 1431   | 4  | US-09-316-083-2    |
| 18         | 44    | 1.2         | 6124   | 4  | US-08-213-419B-3   |
| 19         | 43.8  | 1.2         | 688    | 4  | US-08-998-416-972  |
| 20         | 43.6  | 1.2         | 6124   | 4  | US-08-213-419B-3   |
| 21         | 43.6  | 1.2         | 168575 | 4  | US-09-426-290-1    |
| 22         | 43.4  | 1.2         | 240    | 1  | US-08-628-417-6    |
| 23         | 43.4  | 1.2         | 1368   | 3  | US-08-874-563-5    |
| 24         | 43.4  | 1.2         | 1368   | 3  | US-08-577-483-14   |
| 25         | 43.4  | 1.2         | 3216   | 2  | US-08-828-007-1    |
| 26         | 43.4  | 1.2         | 4253   | 3  | US-08-577-483-7    |
| 27         | 43.4  | 1.2         | 6265   | 4  | US-09-129-112-3    |

|    |      |     |      |   |                    |                    |
|----|------|-----|------|---|--------------------|--------------------|
| 28 | 43.2 | 1.2 | 2447 | 2 | US-09-014-969-14   | Sequence 14, Appl  |
| 29 | 43.2 | 1.2 | 5140 | 4 | US-09-333-214-2    | Sequence 2, Appl   |
| 30 | 43   | 1.2 | 732  | 4 | US-08-998-416-1036 | Sequence 1036, Ap  |
| 31 | 43   | 1.2 | 782  | 4 | US-09-007-119-15   | Sequence 15, Appl  |
| 32 | 42.8 | 1.2 | 1186 | 2 | US-08-731-722-5    | Sequence 5, Appl1  |
| 33 | 42.8 | 1.2 | 1511 | 1 | US-07-991-867B-8   | Sequence 8, Appl1  |
| 34 | 42.8 | 1.2 | 1511 | 1 | US-08-107-755A-8   | Sequence 8, Appl1  |
| 35 | 42.8 | 1.2 | 1511 | 2 | US-08-544-332-8    | Sequence 8, Appl1  |
| 36 | 42.8 | 1.2 | 1511 | 4 | US-09-370-861A-8   | Sequence 8, Appl1  |
| 37 | 42.8 | 1.2 | 3095 | 6 | 5231168-1          | Sequence 8, Appl1  |
| 38 | 42.8 | 1.2 | 6768 | 1 | US-08-107-755A-1   | Patent No. 5231168 |
| 39 | 42.8 | 1.2 | 8457 | 1 | US-07-991-867B-1   | Sequence 1, Appl1  |
| 40 | 42.8 | 1.2 | 8457 | 2 | US-08-544-332-1    | Sequence 1, Appl1  |
| 41 | 42.8 | 1.2 | 8457 | 4 | US-09-370-861A-1   | Sequence 1, Appl1  |
| 42 | 42.6 | 1.2 | 724  | 4 | US-08-998-416-683  | Sequence 683, App  |
| 43 | 42.6 | 1.2 | 1375 | 2 | US-08-736-722A-1   | Sequence 1, Appl1  |
| 44 | 42.6 | 1.2 | 1375 | 3 | US-09-221-114-1    | Sequence 1, Appl1  |
| 45 | 42.4 | 1.2 | 4254 | 2 | US-08-443-639-7    | Sequence 7, Appl1  |

#### ALIGNMENTS

RESULT 1  
US-08-998-416-288  
; Sequence 288, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Reischung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII  
; NUMBER OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NO. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: NO. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 288:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 837 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: PAG1241RP



```

RESULT 4
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
;
; GENERAL INFORMATION:
;
; APPLICANT: Delcive, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasmidogen Activator
; NUMBER OF SEQUENCES: 39

```

|    |     |   |      |
|----|-----|---|------|
| Qy | 515 | TTGTGCTTAATTCGGTACGTCACATATTCATTAATTCACATTAATTAACACATTC   | 574  |
| Dd | 112 | TTTTTAATTAATAATTTAATTAATTAATTTATATATTAATTAATTTAATTAATA    | 171  |
| Qy | 575 | TAAAGCGCTGTTATTGAGTGAATTTGCAATTAATCACAATAAATGATGCTATTAT   | 634  |
| Dd | 172 | TTTTTAATTAATAATTTAATTAATAATTTATTAATTAATAATTTAATTAATA      | 231  |
| Qy | 635 | ACTGCCTATTGATACACTTCACGCTCATACAGATGAATTAATTTGAGGTACACC    | 694  |
| Dd | 232 | TTTAAATTAATAATTTAATTAATAATTTATTAATAATTAATAATTTAATTAATA    | 291  |
| Qy | 695 | TTAATTCAGGTAATTTGCTATTAAATTAATTAATCTGAATTAATTAATGATGCTT   | 754  |
| Dd | 292 | TTTAAATTAATAATGTTAATAATTAATTAATTAATTAATAATTTAATTAATA      | 351  |
| Qy | 755 | GTGCTATTTTTA---TGTTGCTTCATTTTAAATGACGGGAGCTGTGCATTCATAT   | 811  |
| Dd | 352 | TTTTTAATTAATAATGTTAATAATTAATTAATTAATTAATAATAATTAATA       | 411  |
| Qy | 812 | TTTATGATGACACATCTTGATGAGATTAATTAAGATATGTTAAATGACAGCGGTT   | 871  |
| Dd | 412 | TTTTTAATTAATAATGTTAATAATTAATTAATTAATTAATTAATAATAATTAATA   | 471  |
| Qy | 872 | GCGTCTAATTTTAAATTCATTAATAATTAATAATTAATTAATTTGCTCTTT       | 931  |
| Dd | 472 | ATAATTAACATATTTAATTAATAAGATATTTAATTAACATATTTTAATAATTAAGT  | 531  |
| Qy | 932 | TATATGCTGCTTTAAAGAGGTAGATGACCTAAAGCGCCATAAATATGCGCTAAT    | 991  |
| Dd | 532 | ATAATTAACATATTTTATATTAATTCATATTTTAAATTAAGTTAAAGACGAGAAAAT | 591  |
| Qy | 992 | GCCATTTGCTAATTAATTCACCTCAAGATACACTATTTGCCAATTTGACA        | 1039 |







```

:
:   REGISTRATION NUMBER: 38,241
:   REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 919-541-8587
:   TELEFAX: 919-541-8689
:   INFORMATION FOR SEQ ID NO: 1137:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 636 base pairs
:     TYPE: nucleic acid
:     STRANDEDNESS: single
:     TOPOLOGY: linear
:   MOLECULE TYPE: DNA (genomic)
:   ORIGINAL SOURCE:
:   ORGANISM: PAG1692RP
:   US-08-998-416-1137
:
: Query Match      1.3%; Score 48.4; DB 4; Length 636;
: Best Local Similarity 44.7%; Pred. No. 0.0034;
: Matches 234; Conservative 0; Mismatches 286; Indels 4; Gaps 1;
:
: QY 523 TTTAATTCGTTACTGCTCATTCATTCTATTATTCACATTTAAATTAACAGCTTCAAAAGC 582
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 75 TTTAAACATATTATTATTCATTATTAAATTAATTAATTAATTAATTAATTAATTAATTA 134
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 583 TGTATTTGATGAATATTCGAATTTCA---CATTAATTAATTAATTAATTAATTAATTA 638
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 135 TAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 194
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 639 GCTGTATTTGATCAACTTTCATGCTCTATACATGTAATTAATTAATTAATTAATTAAT 698
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 195 TATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 254
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 699 TTCAAGGATATTCGCTATTATTATTATTAATTAATTAATTAATTAATTAATTAATTAAT 758
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 255 TTGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 314
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 759 TTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 818
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 315 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 374
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 819 ATGACAAACATCTTTGATGAAGTATTTAAGATTAATTAATTAATTAATTAATTAATTA 878
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 375 TATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 434
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 879 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 938
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 435 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 494
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 939 GTTCTTTTAAAGAGTAGATGACCTTAAGGTCGCTAATTAATTAATTAATTAATTAATTA 998
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 495 TTTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 554
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 999 CTATTAATTCACCTCAAGATACACTATTGGCAATTAATTAATTAATTAATTAATTA 1042
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 555 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 598
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:
: RESULT 11
: US-09-426-290-1
:   Sequence 1, Application US/09426290
:   Patent No. 6410712
:   GENERAL INFORMATION:
:   APPLICANT: Berglund, Ran Olofsdotter
:   TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
:   FILE REFERENCE: 2345,2001-000
:   CURRENT APPLICATION NUMBER: US/09/426,290
:   CURRENT FILING DATE: 1999-10-25
:   NUMBER OF SEQ ID NOS: 24
:   SOFTWARE: FastSeq for Windows Version 4.0
:   SEQ ID NO 1
:   LENGTH: 168575
:   TYPE: DNA

```

```

:
:   ORGANISM: Homo Sapiens
:   FEATURE:
:   NAME/KEY: CDS
:   LOCATION: (21181)...(21403)
:   NAME/KEY: CDS
:   LOCATION: (95252)...(95430)
:   NAME/KEY: CDS
:   LOCATION: (101753)...(101996)
:   NAME/KEY: CDS
:   LOCATION: (110324)...(110439)
:   NAME/KEY: CDS
:   LOCATION: (124058)...(124278)
:   NAME/KEY: CDS
:   LOCATION: (127009)...(127130)
:   NAME/KEY: CDS
:   LOCATION: (128910)...(129139)
:   US-09-426-290-1
:
: Query Match      1.3%; Score 48; DB 4; Length 168575;
: Best Local Similarity 46.7%; Pred. No. 0.049;
: Matches 185; Conservative 0; Mismatches 210; Indels 1; Gaps 1;
:
: QY 542 TATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 601
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 108944 TATATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109003
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 602 CGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 109004 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109063
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 661 TGTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 109064 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109123
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 721 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 109124 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109183
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 781 TTTAATGAGCGTAGCTGTCATTAATTTTATTAATTAATTAATTAATTAATTAATTA 840
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 109184 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109243
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 841 ATTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 109244 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109303
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   QY 901 AAATCAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 936
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:   DB 109304 ATGTAGGAAATCGAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109339
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
:
: RESULT 12
: US-07-867-106-2/c
:   Sequence 2, Application US/07867106
:   Patent No. 5389526
:   GENERAL INFORMATION:
:   APPLICANT: Slade, Martin B
:   APPLICANT: Chang, Andy C M
:   APPLICANT: Williams, Keith L
:   TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
:   TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
:   NUMBER OF SEQUENCES: 19
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & NO. 5389526r15
:   STREET: One Liberty Place 46th Floor
:   CITY: Philadelphia
:   STATE: PA
:   COUNTRY: USA
:   ZIP: 19103
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS

```



```
US-08-998-416-937
; Sequence 937, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38, 241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 937:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1581RP
US-08-998-416-937

Query Match 1.2%; Score 44.8; DB 4; Length 665;
Best Local Similarity 47.5%; Pred. No. 0.03;
Matches 206; Conservative 0; Mismatches 217; Indels 11; Gaps 2;

QY 527 ATTCGTTACTGTCATATTCATTAATTAATTAACAGTTCAAAAAGCGTGT 586
DB 223 ATTTTAAATGTTAATAAATAAATAAGAAATTAATTAATTAATTAATTAAT 282
QY 587 TATTGATGAATTCGAATTAATCAATTAATTAATTAATTAATTAATTAATTA 646
DB 283 TATTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 342
QY 647 GGTATCACTTCATGCTCTTAATACATGAATATAT--TTGAGTTAGACCTTAATTCGAAG 704
DB 343 AATATTAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 402
QY 705 GTAATTTGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 755
DB 403 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 462
QY 756 TGGTATTTTATGTTTGTTCATTTTAAATGACGGTGTGTCATTTCAATTTT 815
```

```
DB 463 TAATAGTCTACCCCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 522
QY 816 ATGATGACACAACTCTTTGATGAAGTATTTAAGATATTTGATGCGGTTGCT 875
DB 523 ATTAAGAAATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 582
QY 876 GTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 935
DB 583 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 642
QY 936 ACTGTTCTTTTAAA 949
DB 643 AATTTTGATTAATA 656

RESULT 15
US-09-419-459-1/c
; Sequence 1, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2110
; TYPE: DNA
; ORGANISM: Pityomyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1749)
US-09-419-459-1

Query Match 1.2%; Score 44.8; DB 4; Length 2110;
Best Local Similarity 54.4%; Pred. No. 0.05;
Matches 111; Conservative 0; Mismatches 92; Indels 1; Gaps 1;

QY 748 TTGCTTTGCTGTTATTTTATGTTTGTTCATTTTAAATGACGGTGCATTTCA 807
DB 2105 TTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 2046
QY 808 TATTTTATGATGACACATCTTTGATGAAGTATTAAGATTTGTAATGATGAGG 867
DB 2045 ATTTTATATATCTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1986
QY 868 GTTTCGCTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 927
DB 1985 TTTTAATTAATTTCTTTTGTGCTTTA-AGTAAAGAAATTAATTAATTAAT 1927
QY 928 CTTTATATAGTGTCTTTTAAAGA 951
DB 1926 CATTCATTAATTAATTTTAAATA 1903

Search completed: March 2, 2003, 01:09:02
Job time : 1426 secs
```





```
22 70.6 2.0 436 25 US-09-654-617-340109 Sequence 340109,
23 -70.6 2.0 436 27 US-09-684-016-340109 Sequence 340109,
24 70.6 2.0 436 58 US-60-145-485-6606 Sequence 6606, Ap
c 25 70.4 2.0 510 17 US-09-304-517A-249382 Sequence 249382,
c 26 70.4 2.0 510 17 US-09-371-145A-249382 Sequence 249382,
c 27 70.4 2.0 510 18 US-09-440-687-8907 Sequence 8907, Ap
c 28 70.4 2.0 510 25 US-09-654-617-62906 Sequence 62906, A
c 29 70.4 2.0 510 27 US-09-684-016-62906 Sequence 62906, A
c 30 70.4 2.0 510 37 US-09-985-678-249382 Sequence 249382,
c 31 70.4 2.0 510 58 US-60-144-084-39263 Sequence 39263, A
c 32 69.6 1.9 824 64 US-60-207-458-18907 Sequence 18907, A
c 33 69.4 1.9 599 63 US-09-873-402A-79376 Sequence 79376, A
c 34 69.4 1.9 599 64 US-60-209-830-51127 Sequence 51127, A
c 35 69.4 1.9 8093 41 US-10-172-086-26 Sequence 26, Appl
c 36 69.2 1.9 498 25 US-09-654-617-180433 Sequence 180433,
c 37 69.2 1.9 498 27 US-09-806-708A-22 Sequence 22, Appl
c 38 69.2 1.9 700 19 US-09-527-433-909 Sequence 909, App
c 39 69 1.9 805 56 US-60-126-265-127 Sequence 127, App
c 40 69 1.9 504 42 US-10-221-613-270 Sequence 270, App
c 41 68.8 1.9 576 26 US-09-666-355A-5638 Sequence 5638, A
c 42 68.8 1.9 799 18 US-09-404-520-13067 Sequence 13067, A
c 43 68.4 1.9 405 23 US-09-619-643-14685 Sequence 14685, A
c 44 68.4 1.9 612 34 US-09-902-540-1357 Sequence 1357, Ap
c 45 68.4 1.9
```

## ALIGNMENTS

```
RESULT 1
US-09-915-706a-1
: Sequence 1, Application US/09915706A
: GENERAL INFORMATION:
: APPLICANT: NELSON, DAVID R.
: TITLE OF INVENTION: A LIVE, AVIRULENT STRAIN OF V. ANGUILLARUM THAT
: PROTECTS FISH AGAINST INFECTION BY VIRULENT V.
: TITLE OF INVENTION: ANGUILLARUM
: FILE REFERENCE: 5112
: CURRENT APPLICATION NUMBER: US/09/915, 706A
: CURRENT FILING DATE: 2001-07-26
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 1
: LENGTH: 3588
: TYPE: DNA
: ORGANISM: Vibrio anguillarum
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: (3572)
: OTHER INFORMATION: a, t, c, g, other or unknown
US-09-915-706a-1
```

```
Query Match 100.0%; Score 3587; DB 34; Length 3588;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GTGCACCTATTGACATGATGCGTACATGATGACATCCTGCTTGGTAACAAGCGT 60
Db 1 GTGCACCTATTGACATGATGCGTACATGATGACATCCTGCTTGGTAACAAGCGT 60
QY 61 TGTATTAAGCTTGTGCGGTTTCATCAAGTTGAACACAAATCTCATGATTTTCCACTT 120
Db 61 TGTATTAAGCTTGTGCGGTTTCATCAAGTTGAACACAAATCTCATGATTTTCCACTT 120
QY 121 CCGGAAGGGAAGTGAAGTAATACCTTTTGATGATCAGCCTTTCTAGAGCTTTTCAATG 180
Db 121 CCGGAAGGGAAGTGAAGTAATACCTTTTGATGATCAGCCTTTCTAGAGCTTTTCAATG 180
QY 181 ATCTTTTCGCTTACGTTTGAAGAAATCTGCAGCAGCTGTTGTAATTCACAAGCTA 240
Db 181 ATCTTTTCGCTTACGTTTGAAGAAATCTGCAGCAGCTGTTGTAATTCACAAGCTA 240
QY 241 AGTGATCCAAATATCTCTATTGTAATTAAGTAAGTGTGCTTGTCTTGTCTATCTCTGTG 300
```

```
|||||
Db 241 AGTGATCCAAATATCTCTATTGTAATTAAGTAAGTGTGCTTGTCTTGTCTATCTCTGTG 300
QY 301 AATTCGAGAGCTTACATATATACCTGAAGAAAAAGCTTTCCAGAAATCTAATTCGTAAGAC 360
|||||
Db 301 AATTCGAGAGCTTACATATATACCTGAAGAAAAAGCTTTCCAGAAATCTAATTCGTAAGAC 360
QY 361 ACACAACAGCTTACCTAGAGTTTGGTATGATTCATGATTTGGCCGATGGAAGC 420
Db 361 ACACAACAGCTTACCTAGAGTTTGGTATGATTCATGATTTGGCCGATGGAAGC 420
QY 421 GAAACAGACACCCCGGATAGCCTTTCCCTCCGATTAATTCGCGACAAATTAACCTT 480
Db 421 GAAACAGACACCCCGGATAGCCTTTCCCTCCGATTAATTCGCGACAAATTAACCTT 480
QY 481 TTGCTATCTGAACCTTAATGATGAGGAGGAGGTTGGGCTTAATTCGTTCTGCTC 540
Db 481 TTGCTATCTGAACCTTAATGATGAGGAGGAGGTTGGGCTTAATTCGTTCTGCTC 540
QY 541 ATATTCATTAATTCACATTAATTAATTAACAGCTTAAAGGCTGTTAATGATGAATAT 600
Db 541 ATATTCATTAATTCACATTAATTAATTAACAGCTTAAAGGCTGTTAATGATGAATAT 600
QY 601 TCGAATTAATTCACATTAATTAATTAATGATGATTAATTAATTAATTAATTAATTA 660
Db 601 TCGAATTAATTCACATTAATTAATTAATGATGATTAATTAATTAATTAATTAATTA 660
QY 661 TGCCTTATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720
Db 661 TGCCTTATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720
QY 721 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
Db 721 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
QY 781 TTTAATGACGCTGAGCTTGTGATTCATTAATTTTAAATGATGACAAATCTTTGATGAGT 840
Db 781 TTTAATGACGCTGAGCTTGTGATTCATTAATTTTAAATGATGACAAATCTTTGATGAGT 840
QY 841 ATTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900
Db 841 ATTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900
QY 901 AAATCAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960
Db 901 AAATCAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960
QY 961 AACTTAAGGCTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020
Db 961 AACTTAAGGCTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020
QY 1021 AACTTAAGGCTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080
Db 1021 AACTTAAGGCTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080
QY 1081 TTGCTGCAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140
Db 1081 TTGCTGCAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140
QY 1141 TATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
Db 1141 TATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
QY 1201 TAAACAGAGGAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1260
Db 1201 TAAACAGAGGAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1260
QY 1261 TGAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320
Db 1261 TGAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320
QY 1321 TACGTAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
Db 1321 TACGTAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
```

Db 1321 TACGTAATGATTTAATGTGCGCAAACTGCGTCTGAAGCTAAGTCAAAACCCCTAGTG 1380  
Qy 1381 CTGACGAGAGATGCGTTCAAGAGGATGTCTAAATAGTGGAGATTCCTCTGACA 1440  
Db 1381 CTGACGAGAGATGCGTTCAAGAGGATGTCTAAATAGTGGAGATTCCTCTGACA 1440  
Qy 1441 GTTTGTGAGAACAGTTTTCAAAAACAACAGAGATATCGAGCTCATCTGTTGTG 1500  
Db 1441 GTTTGTGAGAACAGTTTTCAAAAACAACAGAGATATCGAGCTCATCTGTTGTG 1500  
Qy 1501 CTGCTCAATTCCTTCTGATACCAATATAGAACTGCTGCGAATAGCTTGATGTTAG 1560  
Db 1501 CTGCTCAATTCCTTCTGATACCAATATAGAACTGCTGCGAATAGCTTGATGTTAG 1560  
Qy 1561 CGGATTAAGTGAAGAACCTGGGATCAGCTCAACCTGTACTACCGTTGAACGCTCA 1620  
Db 1561 CGGATTAAGTGAAGAACCTGGGATCAGCTCAACCTGTACTACCGTTGAACGCTCA 1620  
Qy 1621 AATCTGATGATGATGAAGGCAAGAAAGAGACAGACAGATGCGAAAGTTAAAGCATTTT 1680  
Db 1621 AATCTGATGATGATGAAGGCAAGAAAGAGACAGACAGATGCGAAAGTTAAAGCATTTT 1680  
Qy 1681 TCCAACTAGTCGCGGATAGCGAGAAAGCTCGATTCCTATGCGCGGCTGCTGCACTGC 1740  
Db 1681 TCCAACTAGTCGCGGATAGCGAGAAAGCTCGATTCCTATGCGCGGCTGCTGCACTGC 1740  
Qy 1741 CCTTAGTCGGGGAGATGACGTTTTTTTACTTTTCAAAGTGCAGAGAGAAAAAGCGAAATCA 1800  
Db 1741 CCTTAGTCGGGGAGATGACGTTTTTTTACTTTTCAAAGTGCAGAGAGAAAAAGCGAAATCA 1800  
Qy 1801 GCCAACTGAATCTATGCTTACGACACACGGTGGCGCAAGACGTTTGCATTTCAATTTCA 1860  
Db 1801 GCCAACTGAATCTATGCTTACGACACACGGTGGCGCAAGACGTTTGCATTTCAATTTCA 1860  
Qy 1861 AGATGAAAAAGCCCAACGTTGTGTCAACCATTAGATCGTTTGTCAAGCTTGTGTGACA 1920  
Db 1861 AGATGAAAAAGCCCAACGTTGTGTGTCAACCATTAGATCGTTTGTGTGACA 1920  
Qy 1921 CTAGTGTCAATCTCTAGGAGATCAAGTCAACCTTCGATTTGGCAAGTCACTGCTTA 1980  
Db 1921 CTAGTGTCAATCTCTAGGAGATCAAGTCAACCTTCGATTTGGCAAGTCACTGCTTA 1980  
Qy 1981 CCCCTGTGAAAAAGCTTTGGTTCATCTAAGTGAATTAAGTATAGCAACCGCAAGCGAGG 2040  
Db 1981 CCCCTGTGAAAAAGCTTTGGTTCATCTAAGTGAATTAAGTATAGCAACCGCAAGCGAGG 2040  
Qy 2041 CCAAGACAGTAGAGCAAGAGGTTGCCAAAGTTCAAGTTTCTGAAGGGAGCTGCCAAGCC 2100  
Db 2041 CCAAGACAGTAGAGCAAGAGGTTGCCAAAGTTCAAGTTTCTGAAGGGAGCTGCCAAGCC 2100  
Qy 2101 AATAGATACAAAACATATAGAGGATATACGATGCGATCAGACAGGCTCAGACGCTAA 2160  
Db 2101 AATAGATACAAAACATATAGAGGATATACGATGCGATCAGACAGGCTCAGACGCTAA 2160  
Qy 2161 GCCAACACTTACAGCAGAGAAACCTCTCTGAACCTGGGTAAATTTAAACAAATATGAACGAG 2220  
Db 2161 GCCAACACTTACAGCAGAGAAACCTCTCTGAACCTGGGTAAATTTAAACAAATATGAACGAG 2220  
Qy 2221 ACTTAGCTTTCATTTTGTGAGAGAGTCTCTGATTAATTTTGGCAGAGCGAACCGCATTA 2280  
Db 2221 ACTTAGCTTTCATTTTGTGAGAGAGTCTCTGATTAATTTTGGCAGAGCGAACCGCATTA 2280  
Qy 2281 GCGCAATTTTCATTTTGTGAGAGAGTCTCTGATTAATTTTGGCAGAGCGAACCGCATTA 2340  
Db 2281 GCGCAATTTTCATTTTGTGAGAGAGTCTCTGATTAATTTTGGCAGAGCGAACCGCATTA 2340  
Qy 2341 TGCTGCGAGAAATGATGCGAGCAAAACGGTGAAGCTCTTATGACATTTTAAATCCG 2400  
Db 2341 TGCTGCGAGAAATGATGCGAGCAAAACGGTGAAGCTCTTATGACATTTTAAATCCG 2400  
Qy 2401 CCGGATTTGAATCATCTGATCAGTCTGTTTGTGCGGAGGAGTGAATCAACCGTGGGCA 2460  
Db 2401 CCGGATTTGAATCATCTGATCAGTCTGTTTGTGCGGAGGAGTGAATCAACCGTGGGCA 2460

Qy 2461 TTGAAGCCCCCAACACCTCAACGGAAGCCTTCGTTTGGATCCGGAAGTTGAG 2520  
Db 2461 TTGAAGCCCCCAACACCTCAACGGAAGCCTTCGTTTGGATCCGGAAGTTGAG 2520  
Qy 2521 AGCATGTATCTCAGACTTCCCTGTAGATACCCAATCTAAGCAAGATCAAAAACACAAAT 2580  
Db 2521 AGCATGTATCTCAGACTTCCCTGTAGATACCCAATCTAAGCAAGATCAAAAACACAAAT 2580  
Qy 2581 CATCCGCTAGCTGGGCTCTGAGTTGGTAATTTGTTTAAAAAATTAAGAAAAATCATGGC 2640  
Db 2581 CATCCGCTAGCTGGGCTCTGAGTTGGTAATTTGTTTAAAAAATTAAGAAAAATCATGGC 2640  
Qy 2641 AAGTATTTACATGCGGTGAAGCGGCTTCAAGTTGAGGGCGACGACTATCGCTAGCT 2700  
Db 2641 AAGTATTTACATGCGGTGAAGCGGCTTCAAGTTGAGGGCGACGACTATCGCTAGCT 2700  
Qy 2701 AGAAGCGCTGAAGTAAAAATGACGGTGTGGTTGCAATCAACCTTACTTTGGGGTGG 2760  
Db 2701 AGAAGCGCTGAAGTAAAAATGACGGTGTGGTTGCAATCAACCTTACTTTGGGGTGG 2760  
Qy 2761 CGCTCGTACGTTGCTATGACATCGGTAAAGCGCAATGCGGATTCAGGCAATGTTGG 2820  
Db 2761 CGCTCGTACGTTGCTATGACATCGGTAAAGCGCAATGCGGATTCAGGCAATGTTGG 2820  
Qy 2821 CGTAAAGCAAGTTAGCGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880  
Db 2821 CGTAAAGCAAGTTAGCGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880  
Qy 2881 TTTATTTCAACCCAGCTTAAAGACGCTTAAACCTGTTGAGTTGATTTACTAAGCCTTCTAA 2940  
Db 2881 TTTATTTCAACCCAGCTTAAAGACGCTTAAACCTGTTGAGTTGATTTACTAAGCCTTCTAA 2940  
Qy 2941 CGATGCTCAAGGTGACAGACGTTTACTTCCAAAGTTAACTGAAAAAACAGCTTTAGTTTC 3000  
Db 2941 CGATGCTCAAGGTGACAGACGTTTACTTCCAAAGTTAACTGAAAAAACAGCTTTAGTTTC 3000  
Qy 3001 TTTCAAGGTGAGCGGGAGTACGCGATCTCAACCGTACGAGAGGCTATCTCTTTTCAAC 3060  
Db 3001 TTTCAAGGTGAGCGGGAGTACGCGATCTCAACCGTACGAGAGGCTATCTCTTTTCAAC 3060  
Qy 3061 TTTCTATTTCTCAGAGCATCATATGAGAAAGAGTGTGAATACAAAGCGGTGTGT 3120  
Db 3061 TTTCTATTTCTCAGAGCATCATATGAGAAAGAGTGTGAATACAAAGCGGTGTGT 3120  
Qy 3121 TGTGACTTACGACCTTACCGACCGGGAATGACTTCTGTGAATTAATTTCTTCAATTAGAC 3180  
Db 3121 TGTGACTTACGACCTTACCGACCGGGAATGACTTCTGTGAATTAATTTCTTCAATTAGAC 3180  
Qy 3181 ATGCGCAGTTAATTTGGCATGCTATTTCAATGAATATCTCAATTTTAAAGACCGTTATGGC 3240  
Db 3181 ATGCGCAGTTAATTTGGCATGCTATTTCAATGAATATCTCAATTTTAAAGACCGTTATGGC 3240  
Qy 3241 ATTGAACCTCAACATATAGGCGCTTACTAAGAACCGTGTGACATCACTATGACGTTGA 3300  
Db 3241 ATTGAACCTCAACATATAGGCGCTTACTAAGAACCGTGTGACATCACTATGACGTTGA 3300  
Qy 3301 AACGAATGCGCGCTTAAAGAGAAAGAGCTGCTTTGTGTTGGCTCATTTGGCAGCTT 3360  
Db 3301 AACGAATGCGCGCTTAAAGAGAAAGAGCTGCTTTGTGTTGGCTCATTTGGCAGCTT 3360  
Qy 3361 TTTGAGGACACAAACACAGATCAGAAAAAGTTGATTTGAAGAGGAGAGGTTCAAGGGAT 3420  
Db 3361 TTTGAGGACACAAACACAGATCAGAAAAAGTTGATTTGAAGAGGAGAGGTTCAAGGGAT 3420  
Qy 3421 CGATTAAGACAACTTTCGATACAGTATGATGAGGCAAAATTCACCGCGTCTTTCGTAAGGT 3480  
Db 3421 CGATTAAGACAACTTTCGATACAGTATGATGAGGCAAAATTCACCGCGTCTTTCGTAAGGT 3480  
Qy 3481 TGATTAACAAAGCTTTCATATGATGATACGCAATTTGAAGTGAAGCTTGAAGCTTCCGTTGAT 3540  
Db 3481 TGATTAACAAAGCTTTCATATGATGATACGCAATTTGAAGTGAAGCTTGAAGCTTCCGTTGAT 3540

[illegible]

```

RESULT 2
US-09-346-965A-909
: Sequence 909: Application US/09346965A
: GENERAL INFORMATION:
:   APPLICANT: Wing, Rod A.
:   APPLICANT: Dean, Ralph A.
:   TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ORYZA SATIVA AND USES THEREOF
:   FILE REFERENCE: S-30669A
:   CURRENT APPLICATION NUMBER: US/09/346, 965A
:   CURRENT FILING DATE: 1998-06-30
:   EARLIER APPLICATION NUMBER: 60/091,404
:   EARLIER FILING DATE: 1998-06-30
:   NUMBER OF SEQ ID NOS: 1945
:   SOFTWARE: PatentIn Ver. 2.1
:   SEQ ID NO 909
:   LENGTH: 714
:   TYPE: DNA
:   ORGANISM: Oryza sativa
:   FEATURE:
:     NAME/KEY: source
:     LOCATION: (1)..(714)
:     OTHER INFORMATION: /clone="nbx0003bD01.g"
:   FEATURE:
:     NAME/KEY: misc_feature
:     LOCATION: (1)..(187)
:     OTHER INFORMATION: High quality region
:   FEATURE:
:     NAME/KEY: misc_feature
:     LOCATION: (1)..(714)
:   OTHER INFORMATION: n = a, t, c or g
US-09-346-965A-909

```

|                           |       |                   |           |            |
|---------------------------|-------|-------------------|-----------|------------|
| Query Match               | 2.2%  | Score 78.8        | DB 17     | Length 714 |
| Best Local Similarity     | 48.9% | Pred. No. 1.9e-06 |           |            |
| Matches 209; Conservative | 0;    | Mismatches 218;   | Indels 0; | Gaps 0;    |

[illegible]

```

RESULT 3
US-09-404-520-13069/c
; Sequence 13069, Application US/09404520
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Ghodssi, Azita
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: McIninch, James
; APPLICANT: Timberlake, William E.
; APPLICANT: Yu, Jaehyuk
; TITLE OF INVENTION: Emericeia nidulans Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-10(15498)A
; CURRENT APPLICATION NUMBER: US/09/404,520
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ. ID NOS.: 44345
; SEQ ID NO 13069
; LENGTH: 948
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
; FEATURE:
; OTHER INFORMATION: unsure at all n locations
US-09-404-520-13069

```

|                           |       |                   |          |            |
|---------------------------|-------|-------------------|----------|------------|
| Query Match               | 2.1%  | Score 76.4        | DB 18    | Length 948 |
| Best Local Similarity     | 48.2% | Pred. No. 7.1e-06 |          |            |
| Matches 209; Conservative | 0     | Mismatches 225;   | Indels 0 | Gaps 0     |

|    |     |  |      |
|----|-----|--|------|
| OY | 515 | TTGGTGCTTAATCGTACTGCGCATATTCATTTACATTAATCAGTAAATAAACAACATTC    | 574  |
| Db | 611 | TTTAAATTTTTTTTTTTTTTTTTTTTATTATTATATANTTTTTTTTTTTTTTTTTTTTAA   | 5523 |
| OY | 575 | TAAAGCGCTGTTTATGSGATGATATCGAATAATTGACATTAATTAATGATCTATTAT      | 6343 |
| Db | 551 | TATATTTTTTTTTTTTTTTTTTTTTTTATTTTATATATAATTTTTTTTTTTTTTTTTAT    | 4922 |
| OY | 635 | ACTGCTGATTTGGATCAACTTTCATGCTGTATACATGTAATATTTGGAGTTAGACC       | 6944 |
| Db | 491 | TTTTTAAATTTTTTTTTTTTTTTATATTTTATTTTTTTCGTTTTTTTAAATTTTTTATTTT  | 4322 |
| OY | 695 | TTAATCAAGGTAATTTGCTCATTTAAATATATTATCGAATAATGTAACGATTCCTT       | 7544 |
| Db | 431 | TTTATTATTAATTTTTTTTTTAAATATTATTATTATTTTTCATATTTTTTTTTTATTTTFA  | 3727 |
| OY | 755 | GGTGTTATTTTANGTTGTTCACTTTTAAAGCGGTGACCTTGCCATTCATATTTT         | 8144 |
| Db | 371 | TTTTTTTTTTTTTTTTTTTTTATTTTATTTTATTTTATTTTNNATTTTAAATATTTTTTTT  | 3122 |
| OY | 815 | TATGATGACAACTCTTGATGAGATATTTAAGATATTGTTAAATGATGAGGGCTTGGC      | 8744 |
| Db | 311 | TTTTTTTTTTTATAATTTTTTTTTTTTTTTTTTTTTTTTATTATTTTTTTTTTTTTTTTTT  | 2522 |
| OY | 875 | TGTATTTTTTATATTAATCAATAAATAAACAATATATGTTATTTTGTGCTTTTAA        | 9344 |
| Db | 251 | TATTTTTTTTTTTTTTTTATTTTATTTTTTTTTTTTTTTTATATTTTATTTTTTTTTTTTTT | 1922 |
| OY | 935 | TAGTGTCTTTAA 948   |      |
| Db | 191 | TTTTATTTTTTTTAA 178  |      |

```

RESULT 4
US-09-666-355A-6574
; Sequence 6574, Application US/09666355A
; GENERAL INFORMATION:
; APPLICANT: Fincher, Karen L.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Xiao, Jinhua
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51462)B
; CURRENT APPLICATION NUMBER: US/09/666,355A
; CURRENT FILING DATE: 2000-09-20

```

: PRIOR APPLICATION NUMBER: US 60/155,005  
 : PRIOR FILING DATE: 1999-09-21  
 : NUMBER OF SEQ ID NOS: 25978  
 : SEQ ID NO 6574  
 : LENGTH: 908  
 : TYPE: DNA  
 : ORGANISM: *Gossypium hirsutum*  
 : FEATURE:  
 : OTHER INFORMATION: Clone ID: uc-gsronu33B039b07b1b  
 : US-09-666-355A-8574

|                       |                 |                    |           |             |
|-----------------------|-----------------|--------------------|-----------|-------------|
| Query Match           | 2.1%            | Score 75;          | DB 26;    | Length 908; |
| Best Local Similarity | 49.7%           | Pred. No. 1.4e-05; |           |             |
| Matches 219;          | Conservative 0; | Mismatches 220;    | Indels 2; | Gaps 1;     |

|    |     |   |     |
|----|-----|---|-----|
| QY | 515 | TTTGGTCTTAATTCGGTTCGCTCATATTCATTAATTCACATTAATAAAGACGTC    | 574 |
| Db | 312 | TTTATGTTATTAATTAATTTATTTTATTTTATTAATTAATTAATTAATTTTA      | 371 |
| QY | 575 | TAAAGCGCTGTTATGATGAATATTCGAATATTCACATTAATAATGATGCTATAT    | 634 |
| Db | 372 | TTATTTATTTTATATTAATTTATTTTATTTTATTTTATTTTATTAATATTTAT     | 431 |
| QY | 635 | ACTGCGTATGCTATACACTTTCATGCGTCATACATGTAATATTTGAGTTAGACC    | 694 |
| Db | 432 | ATAATATTAATTAATTAATTTATTTGATTTTATTTATATTAATTTTATTTTAA     | 491 |
| QY | 695 | TTAATTCAGCTAAATTTGCTCATTTATTAATTCATGCAATAATTCGAATCGATCGTT | 754 |
| Db | 492 | TTAATTTATTTTATTTATTTATTTATTAATTTTACTTTAATTAATTTTATTTAT    | 551 |
| QY | 755 | GTCGTTATTTTATGCTTGTGTTCACTTTTAAATGACGCTGACCTGTCATTCATATTT | 814 |
| Db | 552 | TAAATTTTATTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT    | 609 |
| QY | 815 | TATGATGACACATCTTGGATGAGATTAATTAAGATATGTTAATGCAATGAGGGTTCG | 874 |
| Db | 610 | AATTAATTTTAAATTTTAAATTAATTAATTTATTTATTAATTAATTAATTTATTT   | 669 |
| QY | 875 | TGTAATTTTATTAATTAATCATAATAAATCAACAAATATATGTAATTTTGTCTTTTA | 934 |
| Db | 670 | TATTTTATTAATAATTTTATTAATTTATTTTATTAATTAATTAATTAATTAATTTTA | 729 |
| QY | 935 | TAGCTTCTTTAAAGAGTA  | 995 |
| Db | 730 | TTTATATATATTAATTTATTA                                     | 790 |

```

RESULT 5
US-09-619-643-13271/c
: Sequence 13271, Application US/09619643
: GENERAL INFORMATION:
: APPLICANT: Fisher, Dane K.
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: TITLE OF INVENTION: 38-21(51230)B
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/09/619,643
: CURRENT FILING DATE: 2000-07-19
: NUMBER OF SEQ ID NOS: 32236
: SEQ ID NO 13271
: LENGTH: 529
: TYPE: DNA
: ORGANISM: Zea mays
: OTHER INFORMATION: Clone ID: LRB3150-114-P2-K1-E7
US-09-619-643-13271

```

|                       |                |   |          |            |
|-----------------------|----------------|---|----------|------------|
| Query Match           | 2.1%           | Score 73.8  | DB 23    | Length 529 |
| Best Local Similarity | 48.9%          | Pred. No. 2.2e-05                                       |          |            |
| Matches 223           | Conservative 0 | Mismatches 230  | Indels 3 | Gaps 1     |
| OY                    | 515            | TTTGTCGTTAATTCGTTACGTCATATTCATTAATTAATTCATTAATTAACAGTTC | 574      |            |
|                       |                |   |          |            |

|    |     |  |      |
|----|-----|--|------|
| Dd | 484 | TTTTTAACTCTTTTAAATAATTTTATATTATTTATTTATTTATTAATAAATAATTTTGA      | 4295 |
| Oy | 575 | TAAAAGCGCTGTTTATTTGGATGAATATTCGAATATTCACATTAATAATTTGATGTATTAT    | 6344 |
|    |     |  |      |
| Dd | 424 | TAAATATTTTTTTTTTTTATTTATTTATTTACTATTTTTTTTCTAAATTTTATTTATTTATA   | 3655 |
| Oy | 635 | ACTTGCTGTAATTTGGTATCAACTTTTCATGCCTCATACATGTAATAATTTTGGATTTAGCC   | 6944 |
| Dd | 364 | TTTTATTTTTTTTCTCAAATAATTTATTTATTTACATTTTATTTATTTTATTTATTAATTT    | 3050 |
| Oy | 695 | TTAATTCAGGTAATTTGTCTATTTAATTAATTAATTAATCAGAATAATATG--TAAATCATGTC | 7511 |
| Dd | 304 | TATATTTAAATCAATTTTATTAATTTTTTTTATTTTACATATTTTTTTTATTTATTAATTA    | 2458 |
| Oy | 752 | TTTGTGTTAATTTTATTTTANGTTTGTTCATTTTAAATGACGGTAGCTTGTCATTCATATT    | 8111 |
|    |     |  |      |
| Dd | 244 | TTTTTTTCATTTAATTTTCTTTTATTAATTTTTTTTTCTTTAATTTTAAATTTTTTTTT      | 1855 |
| Oy | 812 | TTTTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGTTT      | 8711 |
| Dd | 184 | TTTTTATTTTCCCTTTTTTTTTTTTTATTTTTCTTTTTTAAATTTTAATTTTTTTTTTTT     | 1255 |
| Oy | 872 | GCGGTATTTTTTATTAATAATCAATPAATAAATCAACAATATATGTATTTTGCTCTT        | 9311 |
| Dd | 124 | TCCGTGTTTTTATATAATAAAAAAATTTTTTAAAAATCCCTTTTTTTTTTTTTTTTTNNATT   | 65   |
| Oy | 932 | TTATAGTCTCTTTTAAAGAGGTGAGATGACCTAAA                              | 967  |
| Dd | 64  | AAATATTTATTTATTTTATTAATTTTNTAATAAAAA                             | 29   |

```

RESULT 6
US-09-654-617-341187/c
; Sequence 341187, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21145097D
; CURRENT APPLICATION NUMBER: US/09/654, 617
; CURRENT FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 341187
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: unsure at all n locations
US-09-654-617-341187

```

|                       |                 |                    |           |             |
|-----------------------|-----------------|--------------------|-----------|-------------|
| Query Match           | 2.1%;           | Score 73.8;        | DB 25;    | Length 529; |
| Best Local Similarity | 48.9%;          | Pred. No. 2.2e-05; |           |             |
| Matches 223;          | Conservative 0; | Mismatches 230;    | Indels 3; | Gaps 1      |

[illegible]



|                       |                 |                    |           |              |
|-----------------------|-----------------|--------------------|-----------|--------------|
| Query Match           | 2.08;           | Score 73.2;        | DB 38;    | Length 1776; |
| Best Local Similarity | 47.28;          | Pred. No. 4.3e-05; |           |              |
| Matches 220;          | Conservative 0; | Mismatches 243;    | Indels 3; | Gaps 1       |

[illegible]

|    |     |  |     |
|----|-----|--|-----|
| Db | 243 | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTAATTATTTTTTTTT | 184 |
| Oy | 763 | TTTTATGTTGTTCATTTTTAATGACGGCTTGCCATCATATTTTTATGATGA    | 822 |
| Db | 183 | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTGTTATTT     | 124 |
| Oy | 823 | CACACTCTTGAGATTTTAAGCATATGTTAATGCATGAGGGCTTGCGTGAATTT  | 882 |
| Db | 123 | TTTTTTTTTTTTTTATTTTTTTTTTTTTTANNNTTTTATTTTTTTATTTTTTTT | 64  |
| Oy | 883 | TTAATTAATCAATAAATAACAACATATAATGTTATTTGTCCTTTTATAGTTC   | 942 |
| Db | 63  | TNTTTTTTATATATAAATTAATATATTTTATTTTATTTTAAAAATTTTAA     | 4   |
| Oy | 943 | TTT TTT 945<br>   <br>3 TTT 1                          |     |
| Db | 3   | TTT 1  |     |

```

RESULT 12
US-60-126-265-3039/C
: Sequence 3039, Application US/60126265
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Aspergillus nidulans Genome Sequence And Uses Thereof
: FILE REFERENCE: 04983.0117/38-21(15484)A
: CURRENT APPLICATION NUMBER: US/60/126,265
: CURRENT FILING DATE: 1999-03-25
: NUMBER OF SEQ ID NOS: 4221
: SEQ ID NO 3039
: LENGTH: 961
: TYPE: DNA
: ORGANISM: Aspergillus nidulans
: FEATURE:
: OTHER INFORMATION: unsure at all n locations
US-60-126-265-3039

```

|                           |        |                    |           |             |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match               | 2.0%   | Score 72.6         | DB 56,    | Length 961; |
| Best Local Similarity     | 48.1%; | Pred. No. 4.8e-05; |           |             |
| Matches 204; Conservative | 0;     | Mismatches 218;    | Indels 2; | Gaps 1;     |

|    |     |   |     |
|----|-----|---|-----|
| OY | 523 | TTTAAATGCTACGCTCATATTCGATTAATTCACGTCTTAAATAACAGCTCTAAAGGC       | 582 |
| Db | 552 | TTTNTTTTAAATTTATATTTTAAATTTTTTTTTTATTTATTAATTAATAAATAATATTTTTAT | 503 |
| OY | 583 | TGTTTATATGATGATATATTCGAAATTTATCCAGATATATATGATGCTTTATTTACTGCTG   | 642 |
| Db | 502 | TTTTTATTTTTTTTTTAAATTTTATTAATTAATTTTTTTTATTTTATTTATTTATTTATTTT  | 445 |
| OY | 643 | TATTCGATCAACCTTTCATGCTCTATACATGTAAATATATTCGAGCTTAGACCTTAATCA    | 702 |
| Db | 444 | ATTTTATTTTATTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATTTATATATTTATTTT  | 385 |
| OY | 703 | AGGTAATTTGCTATTTTATTTATTTATTCGATTAATATGATTAATGCTTTGCTGTTAT      | 762 |
| Db | 384 | TTATATTTATTTTTTTTTTAAATTTTATTTTTTTTATTTTTTTTATTTATTTATTTAT      | 322 |
| OY | 763 | TTTATATGTTGCTTTCATTTTTTAATGAGCGTAGCCTTGCAATCATATTTTATATGATGA    | 822 |
| Db | 324 | TTTTTTTTTAAATTTTAAATATTTTNTTTTTTTTTTTTTTTTTTTCTCTTTTTTTTTT      | 265 |
| OY | 823 | CAACATCCTTGATGAGATTTAGATATTTCTTAATGATGATGAGGCTTGGCTATTTT        | 882 |
| Db | 264 | CTTCTCTNTTTCTATNTTTTTTTTTTTTTTTCTTNNNGGTTTAGATNTTTTTTNTTNT      | 205 |
| OY | 883 | TTATATTAATCATATAAATAACAACATATATGTATTTATGCTCTTTTATATAGTCTC       | 942 |
| Db | 204 | TTNNNTTNTTTTTTTTTTAAATTAATCCCTTCCTTTTTTNTTTTTTNTTTTTTNTTTT      | 145 |
| OY | 943 | TTTTT 946   |     |
| Db | 144 | TTCT 141  |     |

```

RESULT 13
US-10-015-127-2817/c
: Sequence 2817, Application US/10015127
GENERAL INFORMATION:
APPLICANT: Bower, Stanley G.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
TITLE OF INVENTION: Sphingomonas elodea genome sequences and uses thereof
FILE REFERENCE: 38-10(15806)B
CURRENT APPLICATION NUMBER: US/10/015,127
CURRENT FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: US 60/252,455
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 14357
: SEQ ID NO 2817
: LENGTH: 1077
: TYPE: DNA
: ORGANISM: Sphingomonas elodea
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(1077)
: OTHER INFORMATION: unsure at all n locations
US-10-015-127-2817

```

|                       |              |                   |                |             |
|-----------------------|--------------|-------------------|----------------|-------------|
| Query Match           | 2.0%         | Score 71.4        | DB 38          | Length 1077 |
| Best Local Similarity | 46.5%        | Pred. No. 9.1e-05 |                |             |
| Matches 201           | Conservative | 0                 | Mismatches 231 | Indels 0    |
|                       |              |                   |                | Gaps 0      |

|    |     |                  |                  |                             |     |
|----|-----|------------------|------------------|-----------------------------|-----|
| QY | 515 | TTTGGCTTAATTCGTA | CTGCTGCAATTCAT   | TAATTCAGTAAATAAAGCTG        | 574 |
| Db | 723 | TTTTTTTTTTTTTTTT | TTTTTTTTTTTTTTTT | TTTTTTTTTTTTTTTT            | 664 |
| QY | 575 | TAAAGCGCTGTTAT   | TGGATGAATTCGA    | ATTTCACATAAATAATGAGCTATAT   | 634 |
| Db | 663 | TTTTTTTTTTTTTTTT | TTTTTAAATTTTTTT  | TTTTTTTTTTTTTTTTNNNTTATA    | 604 |
| QY | 635 | ACTTCCTATTGGTA   | CACTTCATGCTCAT   | ACATGNAATATTGAGTACAC        | 604 |
| Db | 603 | TTTTTTTTTTTTTTTT | TTTTTTTTTTTTTTTT | TTTTTTTTTTTTTTTTATTTTT      | 544 |
| QY | 695 | TTAATTCAGGTAAT   | TCTCATTTAATTA    | TATATCGAATAATGAATGATGCTT    | 754 |
| Db | 543 | TTTTTTTTTTAAAT   | TTTTTTTTTTTTTT   | TTTTTTATTTTTTTTTTTTTTTA     | 484 |
| QY | 755 | GTGCTATTTTANGT   | TGTTGTCATTTT     | TAATGACGGTGAAGCTTGCAATATTTT | 814 |
| Db | 483 | NNATTTTTTTTTTT   | TTTTTTTTTTTTTTTT | TTTTTATTTTTTTTTTTTTTTTTT    | 424 |
| QY | 815 | TATGATGACACAT    | CTTGATGAGATTA    | TAAATTTGTTAATGATGAGGCTTGGC  | 874 |
| Db | 423 | TTTTTTTAAATTT    | TTTTTTTTTTTANT   | TTTTTTTTTTTTTTTTTTTTTTTTT   | 364 |
| QY | 875 | TGTAATTTTAAAT    | TAAATCAATAA      | AAATGATATGTAATTTGTCCTTTTA   | 934 |
| Db | 363 | TTTAAATTTTTTTT   | TTTTTTTTTTTTTT   | TTTTTTTTTAAATTTTATTTTTTTTTT | 304 |
| QY | 935 | TAGTCTCTTT       | 946              |                             |     |
| Db | 303 | TTTTTTTTTTT      | 292              |                             |     |

```

RESULT 14
US-09-849-526A-11985/C
; Sequence 11985, Application US/09849526A
; GENERAL INFORMATION:
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Shukla, Hridayabhijan
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 16517.250[38-21(51930)B]
; CURRENT APPLICATION NUMBER: US/09/849,526A

```

```
; CURRENT FILING DATE: 2001-05-07
; PRIOR APPLICATION NUMBER: US 60/202,214
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/684,016
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/816,660
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 30131
; SEQ ID NO 11985
; LENGTH: 589
; TYPE: DNA
; ORGANISM: Zea mays subsp. mexicana
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(589)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: uc-zmroteoslnle051f12a1
US-09-849-526A-11985
```

```
Query Match      2.0%; Score 71; DB 32; Length 589;
Best Local Similarity 46.5%; Pred. No. 9.1e-05;
Matches 227; Conservative 0; Mismatches 261; Indels 0; Gaps 0;
```

```
QY 457 ATTAATTCGCTGACAATATACCTTTGCTATCTGAAAGCTTAATGCGAGGAGCGGCTT 516
  || || || || || || || || || || || || || || || || || || || || ||
DB 517 ATTAATTCGCTGACAATATACCTTTGCTATCTGAAAGCTTAATGCGAGGAGCGGCTT 516
  || || || || || || || || || || || || || || || || || || || || ||
QY 517 TGGTCTTAATTCGCTGCTGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTAT 576
  || || || || || || || || || || || || || || || || || || || || ||
DB 457 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 398
  || || || || || || || || || || || || || || || || || || || || ||
QY 577 AAGGCTGCTTATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 636
  || || || || || || || || || || || || || || || || || || || || ||
DB 397 TATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 338
  || || || || || || || || || || || || || || || || || || || || ||
QY 637 TTGCTGATTCGATCACTTCATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTAT 696
  || || || || || || || || || || || || || || || || || || || || ||
DB 337 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 278
  || || || || || || || || || || || || || || || || || || || || ||
QY 697 AATTCAGGTAATTCGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 756
  || || || || || || || || || || || || || || || || || || || || ||
DB 277 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 218
  || || || || || || || || || || || || || || || || || || || || ||
QY 757 GGTATTTATTCGCTGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 816
  || || || || || || || || || || || || || || || || || || || || ||
DB 217 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 158
  || || || || || || || || || || || || || || || || || || || || ||
QY 817 TGATGACACATCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876
  || || || || || || || || || || || || || || || || || || || || ||
DB 157 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 98
  || || || || || || || || || || || || || || || || || || || || ||
QY 877 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 936
  || || || || || || || || || || || || || || || || || || || || ||
DB 97 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 38
  || || || || || || || || || || || || || || || || || || || || ||
QY 937 GTGTTCTT 944
  || || || || || || || || || || || || || || || || || || || || ||
DB 37 GTTTATTT 30
  || || || || || || || || || || || || || || || || || || || || ||
```

```
RESULT 15
US-60-202-214-11985/c
; Sequence 11985, Application US/60202214
; GENERAL INFORMATION:
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Shukla, Hridayabhiranjan
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(51930)A
; CURRENT APPLICATION NUMBER: US/60/202,214
; CURRENT FILING DATE: 2000-05-08
```

```
; NUMBER OF SEQ ID NOS: 29880
; SEQ ID NO 11985
; LENGTH: 589
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: uc-zmroteoslnle051f12a1
US-60-202-214-11985
```

```
Query Match      2.0%; Score 71; DB 64; Length 589;
Best Local Similarity 46.5%; Pred. No. 9.1e-05;
Matches 227; Conservative 0; Mismatches 261; Indels 0; Gaps 0;
```

```
QY 457 ATTAATTCGCTGACAATATACCTTTGCTATCTGAAAGCTTAATGCGAGGAGCGGCTT 516
  || || || || || || || || || || || || || || || || || || || || ||
DB 517 ATTAATTCGCTGACAATATACCTTTGCTATCTGAAAGCTTAATGCGAGGAGCGGCTT 516
  || || || || || || || || || || || || || || || || || || || || ||
QY 517 TGGTCTTAATTCGCTGCTGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTAT 576
  || || || || || || || || || || || || || || || || || || || || ||
DB 457 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 398
  || || || || || || || || || || || || || || || || || || || || ||
QY 577 AAGGCTGCTTATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 636
  || || || || || || || || || || || || || || || || || || || || ||
DB 397 TATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 338
  || || || || || || || || || || || || || || || || || || || || ||
QY 637 TTGCTGATTCGATCACTTCATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTAT 696
  || || || || || || || || || || || || || || || || || || || || ||
DB 337 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 278
  || || || || || || || || || || || || || || || || || || || || ||
QY 697 AATTCAGGTAATTCGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 756
  || || || || || || || || || || || || || || || || || || || || ||
DB 277 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 218
  || || || || || || || || || || || || || || || || || || || || ||
QY 757 GGTATTTATTCGCTGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 816
  || || || || || || || || || || || || || || || || || || || || ||
DB 217 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 158
  || || || || || || || || || || || || || || || || || || || || ||
QY 817 TGATGACACATCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876
  || || || || || || || || || || || || || || || || || || || || ||
DB 157 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 98
  || || || || || || || || || || || || || || || || || || || || ||
QY 877 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 936
  || || || || || || || || || || || || || || || || || || || || ||
DB 97 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 38
  || || || || || || || || || || || || || || || || || || || || ||
QY 937 GTGTTCTT 944
  || || || || || || || || || || || || || || || || || || || || ||
DB 37 GTTTATTT 30
  || || || || || || || || || || || || || || || || || || || || ||
```

```
Search completed: March 2, 2003, 03:23:35
Job time : 7550 secs
```









```

Db 1091 WDRFTKDDMDWTWMTWMDRADRRNAGDADRMADGATWMTATWMTWMTWMTWMD 1032
QY 674 TAATATATTCAGTAGACCTTAATCAAGTAATTTGCTATTTATTAATTAATTCGAA 733
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 1031 TWMKMMWMAAATDPAWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMT 972
QY 734 TAATATGATGATTCGTTGCTGTTATTTATTTGTTTCAATTTTAATGACGCTG 793
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 971 KDKKRDGDDGDKKGGKKKAAKAAKATWMTWMDWMDKDKWMDGAKRDKADDDGAGDK 912
QY 794 AGCTTGACATTCATATTTTATGATGACACATCTTGATGAGTATTTAATGATATG 853
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 911 DDDGKADADDDTGTGTDKDDKDKMDKGTWGDATWMAATDMMWMTWMDADWMTW 852
QY 854 TTAATGATGAGGCTTTCGTGATTTTATTAATTAATTAATTAATTAATTAATTAAT 911
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 851 DAADDDWMDADRRMAMKMDAMWAGARTADRRDMDGDRACKRGARRRRRRKRAADKRD 792
QY 912 ATATGTTATTTTGTCTTTTATAGTCTTTTAAAGAGTAGATGACCTAAAGCTC 971
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 791 AADRDPAATWTTTTRDPTDMDKDRRRKARRRRRTTARAAWMTWMTWMDKMDKTRAD 732
QY 972 GCCCAATATGCGTAATTCGCTGATTAATTCACCTCAAGATACATATTCGCA 1031
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 731 RTWRRMRKRDTRMDADADRDADRRRRDGDADACGKRTGRRRRRRRDRATWTRDA 672
QY 1032 ATGACAAATATGTCACCTGATGAAACATATTAATTAATTAATTAATTAATTAAT 1091
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 671 WMAADAAWTTTDTDTDDMDKDRRRKARRRRRTTARAAWMTWMTWMDKMDKTRAD 612
QY 1092 ATAAATATTTTGTGTTGAAATACCAAGGCTGACCTTTCTTTATTCCTAAAT 1151
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 611 RMDWMAADTWTDAKADKADMDKARAWRRARARARARARARARARARARARARAR 552
QY 1152 ACAGGAATACCAATGACGTAATGACCTTAAGCAATAGCAACCTAACAGAGAG 1211
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 551 AAMWMAATTAATWTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 492
QY 1212 GAACCTATGCTTGTCAACATCAATGACCACTTCTTAACCTCTGAGTGATGAT 1271
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 491 AAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 432
QY 1272 TCGATCTGTGCGTTTATCTTAACGAAAGAGTCTTTGCGCC 1318
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 431 WAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 385
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :

RESULT 5
CNS00H04 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
LOCUS BACR35018 of RPCI-98 library from Drosophila melanogaster (fruit
ACCESSION AL074046.1 GI:4953725
VERSION AL074046.1 GI:4953725
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT
- Web : www.genoscope.cns.fr
- Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazutoyo Osoegawa and

```

```

FEATURES
    source
        1..1101
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="BACR35018"
            /clone_1id="RPCI-98"
            /note="end : 17"
            location/Qualifiers

BASE COUNT 168 a 22 c 92 g 398 t 421 others

ORIGIN
Query Match 2.1%; Score 76.6; DB 17; Length 1101;
Best Local Similarity 34.7%; Pred. No. 5.6e-07;
Matches 191; Conservative 11; Mismatches 348; Indels 0; Gaps 0;

QY 628 TATTATTACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 687
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 48 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 107
QY 688 TTAGACCTTAATCAAGGATTTGCTATTTATTTATTTATTTATTTATTTATTTATTT 747
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 108 TTGTDTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 167
QY 748 TTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 807
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 168 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 227
QY 808 TATTATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 867
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 228 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 287
QY 868 GTTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 927
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 288 NTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 347
QY 928 CTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 987
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 348 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 407
QY 988 AATTGCAATGCTATTAATTCACCTCAAGATACCTATTTGCAATTTGACAAATATG 1047
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 408 NNATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 467
QY 1048 CTTGCTGTAAGCAATTTTATGATGATGATGATGATGATGATGATGATGATGATG 1107
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 468 NTATANNNTNNNTATTTACANNANNAANNNNNNNNNNNNNNNNNNNNNNNNNNN 527
QY 1108 TTGAATTAACCAAGGCTCTAGGCTTTCTTATCTTAATTAATCAAGAAATAGCAT 1167
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 528 TTTAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 587
QY 1168 GAAGTAAAT 1177
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 588 TTTNTTTNTT 597
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :

RESULT 6
CNS009D0/c 781 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR19009 of RPCI-98 library from Drosophila melanogaster (fruit
ACCESSION AL053444.1 GI:4934889
VERSION AL053444.1 GI:4934889
KEYWORDS GSS.

```

| SOURCE                   | ORGANISM   |
|--------------------------|--|
| Drosophila melanogaster. | Drosophila melanogaster.   |
| Drosophila melanogaster  | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |
| REFERENCE                | 1 (bases 1 to 781)   |
| AUTHORS                  | Genoscope.   |
| TITLE                    | Direct Submission  |
| JOURNAL                  | Genoscope - Centre National de Sequencage : Submitted (02-JUN-1999)  |

**COMMENT**

determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oseegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPC1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

source
1. .781
/organism="Drosophila melanogaster"
/db-xref="taxon:7227"
/cclone="BACR19009"
/cclone_1lb="RPC1-98"
/notes="end = TERT"
440 a 71 c 6 g 189 t 75 others
BASE COUNT
ORIGIN

```

|                           |       |                   |          |            |
|---------------------------|-------|-------------------|----------|------------|
| Query Match               | 2.1%  | Score 75.2        | DB 17    | Length 781 |
| Best Local Similarity     | 46.0% | Pred. No. 1.1e-06 |          |            |
| Matches 203; Conservative | 21    | Mismatches 215    | Indels 2 | Gaps 1     |

|    |      |  |     |
|----|------|--|-----|
| QY | .515 | TTTGCGCTGCTAAATCGCTATCGCATATTCATTAATTCACATTTAAATAAGAGTTC         | 574 |
| Db | 439  | TTTATTTTTTTTAAATATATTATTTTTTTTTTTTAAATTTAAATTTAAATTTTA           | 380 |
| QY | 575  | TAAAGCGCTGTTTATGGATGAATATTCGAATAATATACATTAATATGATCGCTATTAT       | 634 |
| Db | 379  | TAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT          | 320 |
| QY | 635  | ACTGCGTATTTGGTATCACATTTTCATGCGCTATATACATGAATATATTTCCAGTTAGACC    | 694 |
| Db | 319  | TTTTTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAAATTTAAAT    | 260 |
| QY | 695  | TTAATTCAGGTAATTTTGCTATTATTAATTTATTCATGATATATATGATCATGATGCTT      | 754 |
| Db | 259  | TAAATTTTAA--TAAATTTTTTTTATTTTATTTATTTATTTAAATTTTTTTTTTTTAAW      | 202 |
| QY | 755  | GTCGTATTTTATGATTTGTTGTTCAATTTTAAATGACGAGCTGTCATCATATTTT          | 814 |
| Db | 201  | TAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT       | 142 |
| QY | 815  | TATATGTACACACATCTTTGATGCAAGTATTTAAGATATTGTTATATGACATGAGGCGTTGCG  | 874 |
| Db | 141  | TWTTTTTATATTTTTTATTTAAATTTTWTATTTTAAATATTAATTTTATTTATTTAAATTTT   | 82  |
| QY | 875  | TGATTTTTTTTTTATTTAAATCATATTAATAAATCAACATATATGTTTATTTGTCGCTTTTA   | 934 |
| Db | 81   | TTTTTTTATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTATTTATTTATTTATTTATTTT | 22  |
| QY | 935  | TAGTGTCTTTTAAAGAGTA  | 955 |
| Db | 21   | TAAATTTATTTATTTATTTATTTA   | 1   |

|            |  |
|------------|--|
| RESULT 7   |  |
| CNS0039C   |  |
| LOCUS      | 1101 bp DNA linear GSS 03-JUN-1995                                       |
| DEFINITION | CNS0039C Drosophila melanogaster genome survey sequence TE3 end of BAC # |
|            | BAOR08K10 of Rpcl-36 library from Drosophila melanogaster (fruit         |
|            | fly), genomic survey sequence.   |
| ACCESSION  | AL0633921  |

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

GSS.  
Drosophila melanogaster.  
Drosophila melanogaster  
Eukaryota: Metazoa: Arthropoda: Hexapoda: Insecta: Pterygota;  
Neoptera: Endopterygota: Diptera: Brachycera; Muscomorpha;  
Ephydroidea: Drosophilidae: Drosophila.  
1 (bases 1 to 1101)

**JOURNAL** submitted (02-JUN-1999) Genoscope - Centre National de Séquençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr) - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

**COMMENT** Determination of this BAC-end sequence was carried out as part of a

melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammoser in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila-bac.htm>.

| FEATURES   | SOURCE | location/Qualifiers                 |
|------------|--------|-------------------------------------|
|            |        | 1. 1101                             |
|            |        | /organism="Drosophila melanogaster" |
|            |        | /db_xref="taxon:7227"               |
|            |        | /clone="BACR08K10"                  |
|            |        | /clone_id="RPC1-98"                 |
|            |        | /note="end : TERT3"                 |
| BASE COUNT |        | 201 a 64 c 131 g 202 t 503 others   |
| ORIGIN     |        |                                     |

|                           |       |                   |          |             |
|---------------------------|-------|-------------------|----------|-------------|
| Query Match               | 2.1%  | Score 74.4        | DB 17    | Length 1101 |
| Best Local Similarity     | 18.3% | Pred. No. 1.8e-06 |          |             |
| Matches 106: Conservative | 241   | Mismatches 233    | Indels 0 | Gaps 0      |

[illegible]

[illegible]

|            |   |            |     |        |                 |
|------------|---|------------|-----|--------|-----------------|
| LOCUS      | CN5035N7  | 576 bp     | DNA | linear | GSS 15-MAY-2000 |
| DEFINITION | Tetradon nigroviridis genome survey sequence PUC-ori end of clone 21A06 of library G from Tetradon nigroviridis, genomic survey sequence. |            |     |        |                 |
| ACCESSION  | AL228940  | GI:7887933 |     |        |                 |
| VERSION    | AL228940.1  |            |     |        |                 |
| KEYWORDS   | GSS: genome survey sequence.  |            |     |        |                 |
| SOURCE     | Tetradon nigroviridis.  |            |     |        |                 |
| ORGANISM   | Tetradon nigroviridis   |            |     |        |                 |

| REFERENCE          | AUTHORS   | TITLE  | JOURNAL     | REFERENCE          | AUTHORS   |
|--------------------|---|--|-------------|--------------------|---|
| 1 (bases 1 to 576) | Roest-Croliulus, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C., Benoit, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J. | Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence | Unpublished | 2 (bases 1 to 576) | Roest-Croliulus, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., |

|           |   |
|-----------|---|
| TITLE     | Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>   |
| JOURNAL   | Unpublished   |
| REFERENCE | 3 (bases 1 to 576)  |
| AUTHORS   | Genoscope.  |
| TITLE     | Direct Submission   |
| JOURNAL   | Submitted (12-APR-2000)   |
| COMMENT   | This sequence is a single read and was generated as part of a large scale clone-end and sequencing project of the <i>Tetraodon nigroviridis</i> genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/tetraodon">http://www.genoscope.cns.fr/tetraodon</a> . |

| FEATURES   | source | location/Qualifiers  |
|------------|--------|--|
|            | 1..576 |  |
|            |        | /organism="Tetraodon nigroviridis"                           |
|            |        | /db_xref="taxon:9983"  |
|            |        | /clone="214A06"  |
|            |        | /clone_lib="G"   |
|            |        | /note="Genoscope sequence ID : C0AG214BA03SP1-end ; PUC-Or1" |
| BASE COUNT | 62 a   | 5 c 18 g 444 t 47 others                                     |
| ORIGIN     |        |  |

|                           |       |  |          |            |
|---------------------------|-------|--|----------|------------|
| Query Match               | 2.1%  | Score 74.2   | DB 17    | Length 576 |
| Best Local Similarity     | 43.7% | Pred. No. 1,9e-06  |          |            |
| Matches 178; Conservative | 27    | Mismatches 202   | Indels 0 | Gaps 0     |
| Oy                        | 542   | TATTCATTTAATCTACTATTAATTAACACGTTCTAAAGCGTGTATTATGGTGAATATT | 601      |            |
| Db                        | 3     | TTTTTTTTTTTTTTTTTTTTGGATTTTGAAAATAAATAAATTAATTAAGAT        | 62       |            |

[illegible]

|            |  |         |     |        |                 |
|------------|--|---------|-----|--------|-----------------|
| LOCUS      | CNS01219   | 1101 bp | DNA | linear | GS5 26-Jul-1999 |
| DEFINITION | Drosophila melanogaster genome survey sequence Sp6 end of BAC<br>BA000824 of DrosBAC library from Drosophila melanogaster (fruit<br>fly), genomic survey sequence. |         |     |        |                 |
| ACCESSION  | AL101595   |         |     |        |                 |
| VERSION    | AL101595.1 GI:5613206  |         |     |        |                 |
| KEYWORDS   | GSS.   |         |     |        |                 |
| SOURCE     | Drosophila melanogaster.   |         |     |        |                 |
| ORGANISM   | Drosophila melanogaster  |         |     |        |                 |

| REFERENCE           | AUTHORS    | TITLE             | JOURNAL  | COMMENT   |
|---------------------|------------|-------------------|--|---|
| 1 (bases 1 to 1101) | Genoscope. | Direct Submission | Submitted (22-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segreifgenoscope.cns.fr - Web : www.genoscope.cns.fr) | Determination of this BAC-end sequence was carried out as part of a |

|          |  |
|----------|--|
| COMMENT  | Determination of this BAC-end sequence was carried out as part of collaboration with the European Drosophila Genome Project (EDGP) - <a href="http://www.edgp.edi.ac.uk">http://www.edgp.edi.ac.uk</a> . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBel0BAC11. |
| FEATURES | Location/Qualifiers  |
| SOURCE   | 1..1101  |

| FEATURES   | source                  | location/Qualifiers |
|------------|-------------------------|---------------------|
|            | 1. 1101                 |                     |
|            | /organism="Drosophila   | melanogaster"       |
|            | /db_xref="taxon:7227"   |                     |
|            | /clone="BACN08024"      |                     |
|            | /clone_1lb="DrosBAC"    |                     |
|            | /plasmid="pBelBAC11"    |                     |
|            | /note="end : SP6"       |                     |
| BASE COUNT | 436 a 142 c 169 g 227 t | 127 others          |
| ORIGIN     |                         |                     |

|                       |       |   |       |                                  |
|-----------------------|-------|---|-------|----------------------------------|
| Query Match           | 2.1%  | Score 73.8  | DB 17 | Length 1101                      |
| Best Local Similarity | 43.5% | Pred. No. 2,5e-06   |       |                                  |
| Matches               | 178   | Conservative  | 38    | Mismatches 190, Indels 3, Gaps 1 |
| OY                    | 543   | ATTCATTTATTCACATATTAAATTAACACATCTTAAAGCGCTTTATTTGATGAGATATTC    | 602   |                                  |
|                       |       |   |       |                                  |
| Db                    | 574   | ATCCCGCATTTTATTAATTAATTAATTCGGCTTTTTCATCTTAATTAATTAATTTATTTT    | 516   |                                  |
| OY                    | 603   | GAAATTTATCACATTAATTAATTTGATGCTATTATTACTGCTGTAATTTGATCAACCTTTCAG | 662   |                                  |

[illegible]

|            |  |
|------------|--|
| RESULT 10  |  |
| CNS00FUH/c | CNS00FUH 996 bp DNA linear GSS 03-JUN-1999   |
| LOCUS      |  |
| DEFINITION | Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR31021 of RPE1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  |
| ACCESSION  |  |
| VERSION    | AL071063   |
| KEYWORDS   | AL071063.1 GI:4951105  |
| SOURCE     | GSS.   |
| ORGANISM   | Drosophila melanogaster.   |
|            | Drosophila melanogaster  |
|            | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |
| REFERENCE  | 1 (bases 1 to 996)   |
| AUTHORS    | Genoscope.   |
| TITLE      | Direct Submission  |
| JOURNAL    | Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  |

**COMMENT**

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library) was prepared by Kazuhiro Oseegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

| FEATURES   | SOURCE | LOCATION/Qualifiers                 |
|------------|--------|-------------------------------------|
| 1.         | 996    |                                     |
|            |        | /organism="Drosophila melanogaster" |
|            |        | /db_xref="taxon:7227"               |
|            |        | /clone="BACR31021"                  |
|            |        | /clone_1lb="RPC1-98"                |
|            |        | /note="end : TET3"                  |
| BASE COUNT | 383 a  | 164 c 81 g 171 t 197 others         |
| ORIGIN     |        |                                     |

|                           |        |                    |           |             |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match               | 2.08;  | Score 73.4;        | DB 17;    | Length 996; |
| Best Local Similarity     | 39.28; | Pred. No. 3.1e-06; |           |             |
| Matches 151; Conservative | 62;    | Mismatches 169;    | Indels 3; | Gaps 1;     |

[illegible]

|            |   |            |     |        |
|------------|---|------------|-----|--------|
| RESULT 11  |   |            |     |        |
| CNS009XJ/c |   |            |     |        |
| LOCUS      | CNS009XJ  | 870 bp     | DNA | linear |
| DEFINITION | Drosophila melanogaster genome survey sequence T7 end of Bac #                                  |            |     |        |
| ACCESSION  | BACR2DH04 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. |            |     |        |
| VERSION    | AI054462  |            |     |        |
| KEYWORDS   | AI054462.1  | GI:4935635 |     |        |
| SOURCE     | GSS.  |            |     |        |
| ORGANISM   | Drosophila melanogaster.  |            |     |        |
|            | Drosophila melanogaster   |            |     |        |
|            | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;                                   |            |     |        |
|            | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;                                      |            |     |        |
|            | Ephydrioidae; Drosophilidae; Drosophila.  |            |     |        |
| REFERENCE  | 1 (bases 1 to 870)  |            |     |        |

JOURNAL  
Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
COMMENT  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuhiro Oosawa and  
Aatón Mammossor in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPcI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source   | 1. .870             |

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR20H04"
/clone_1lp="RPC1-98"
/note="end : 17"
BASE COUNT      404 a      46 c      100 g      155 t      165 others
ORIGIN

```

|                       |                  |                    |           |             |
|-----------------------|------------------|--------------------|-----------|-------------|
| Query Match           | 2.0%;            | Score 73.2;        | DB 17;    | Length 870; |
| Best Local Similarity | 40.6%;           | Pred. No. 3.3e-06; |           |             |
| Matches 164;          | Conservative 33; | Mismatches 207;    | Indels 0; | Gaps 0;     |

[illegible]

|            |   |
|------------|---|
| RESULT     | 12  |
| CNS0012D   |   |
| LOCUS      | CNS0012D  |
| DEFINITION | Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR36013 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.   |
| ACCESSION  | AL074343  |
| VERSION    | GI:4954012  |
| KEYWORDS   | GSS.  |
| SOURCE     | Drosophila melanogaster.  |
| ORGANISM   | Drosophila melanogaster.<br>Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  |
| REFERENCE  | 1 (bases 1 to 991)<br>Genoscope.<br>Direct Submission<br>Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :<br>BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr<br>- Web : www.genoscope.cns.fr)   |
| COMMENT    | Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library">http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library</a> was prepared by Kazutoyo Osoegawa and Aaton Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp. the same strain used for the library p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> . |
| FEATURES   | location/Qualifiers<br>1..991   |
| SOURCE     | /organism="Drosophila melanogaster"<br>/db_xref="taxon:7229"  |

|            |  |      |      |       |            |
|------------|--|------|------|-------|------------|
| BASE COUNT | 176 a  | 95 c | 64 g | 515 t | 141 others |
| ORIGIN     | /clone="BACR36013"<br>/clone_lib="RPC1-98"<br>/note="end : TET3" |      |      |       |            |

|            |   |                                       |
|------------|---|---------------------------------------|
|            | Query Match   | 2.0%; Score 72.8; DB 17; Length 991;  |
|            | Best Local Similarity   | 39.0%; Pred. No. 4.2e-06;             |
|            | Matches 194; Conservative   | 30; Mismatches 274; Indels 0; Gaps 0; |
| QY         | 449 CCTCTCGATTAAATTCGGTGCACATAATATACCTTTGCTATCGAAGCTTAATGCTGAGG     | 508                                   |
| Db         | 57 CCCTTACAAGAAGATCTCKAAAAATTAAAATTTAAATTTCCCTNTAACANNNNNAATKT      | 116                                   |
| QY         | 509 AGCGGCTTGSTGCTTAATTCGTACGCGCATATTCATTAATTCACATTAATAATAA         | 568                                   |
| Db         | 117 TATATACCTGGATTAATTTWTTAAATAACCTCACAAAACCTTWATTAATATCTTTAK       | 176                                   |
| QY         | 569 CAGTTCCTAAAGGCTGTTATTCGATGAATATTCGACATATCAATATATGATGCT          | 628                                   |
| Db         | 177 ACCATTTTTNNTNTNTNTNTNTNTNTNTNTMNCNTNNAAATATWYCWTITTTTWICKTTTT   | 236                                   |
| QY         | 629 ATTAATTCCTGCTGATTCGATGATCAGCTTCATGCTATACATGTAATATATTCGAGT       | 688                                   |
| Db         | 237 TTTTTTYCYCTTTTTCATCAATATAMWTTCMKTTTTTTTTCATATVTVNNNNNNNN        | 296                                   |
| QY         | 689 TAGACCTTAATTCAGGAATTTGCTATTTAATTTATATCTGAATATATGATGAT           | 748                                   |
| Db         | 297 TNNNNNNNNNTNNNTNNNTTYCTTTTTTTTTTTTTTTTTTTTTTTCMTTTTTTTT         | 356                                   |
| QY         | 749 TGCTTTTGATTTATTTATGTTTTCATTTTAAATGAGCGTAGCTGATTCAT              | 808                                   |
| Db         | 357 YTTCTY    | 416                                   |
| QY         | 809 ATTATTATGATGACACATCTTGATGACAGATTTAAGATATTTGTATGACAGAGGG         | 868                                   |
| Db         | 417 CTTTKTT   | 476                                   |
| QY         | 869 TTTCGCTGCTTTTATTAATTAATCATTAATAAATCAACAATATGCTATTTGCTGC         | 928                                   |
| Db         | 477 TTTTCTCTKTT     | 536                                   |
| QY         | 929 TTTTATAGTGTCTTTT  | 946                                   |
| Db         | 537 TTTTTTTTTTTTTTTTTT  | 554                                   |
| RESULT 13  | CNS06PLB  | 1084 bp DNA linear GSS 05-JUL-2001    |
| LOCUS      | T7 end of clone AV0AA015D02 of library AV0AA from strain CBS 379 of |                                       |
| DEFINITION | Saccharomyces exiguus, genomic survey sequence.                     |                                       |
| ACCESSION  | AL409509  |                                       |
| VERSION    | AL409509.1 GI:12177058.   |                                       |
| KEYWORDS   | GSS.  |                                       |
| SOURCE     | Saccharomyces exiguus.  |                                       |
| ORGANISM   | Saccharomyces exiguus   |                                       |
| REFERENCE  | Eukaryota: Fungi: Ascomycota: Saccharomycotina: Saccharomycetes;    |                                       |
| AUTHORS    | Saccharomycetales: Saccharomycetaceae; Saccharomyces.               |                                       |
|            | 1 (bases 1 to 1084)   |                                       |
|            | Soucié,J.L., Aigle,M., Artiguenave,F., Blandin,G.,                  |                                       |
|            | Boiotin-pukuhara,M., Bon,E., Brothier,P., Casaregola,S.,            |                                       |
|            | de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,     |                                       |
|            | Malpertuy,A., Neugeglise,C., Ozler-Kalogenopoulos,O., Potier,S.,    |                                       |
|            | Saurin,W., Tekala,F., Toffano-Nioche,C., Mesolowski-Louvel,M.,      |                                       |
|            | Wincker,P. and Weissenbach,J.                                       |                                       |
|            | Genomic exploration of the hemiascomycetous yeasts: 1. A set of     |                                       |
|            | yeast species for molecular evolution studies                       |                                       |
| JOURNAL    | FEMS Lett. 487 (1), 3-12 (2000)                                     |                                       |
| MEDLINE    | 20584711  |                                       |
| PUBMED     | 11152876  |                                       |
| REFERENCE  | 2 (bases 1 to 1084)   |                                       |
| AUTHORS    | Bon,E., Neugeglise,C., Lepingle,A., Wincker,P., Artiguenave,F.,     |                                       |

| TITLE   | JOURNAL                                    | MEDLINE  | PUBMED    | REFERENCE    | AUTHORS   | TITLE | JOURNAL | COMMENT |
|---|--|--|-----------|--------------|-----------|-------|---------|---------|
| Gallardin,C. and Casaregola,S.                                      |  |  |           |              |           |       |         |         |
| Genomic exploration of the hemiascomycetous yeasts: 6.              |  |  |           |              |           |       |         |         |
| Saccharomyces exiguus   | FEBS Lett.                                 | 487 (1), 42-46   | (2000)    |              |           |       |         |         |
| 20584716  |  |  |           |              |           |       |         |         |
| 1152881   |  |  |           |              |           |       |         |         |
| 3 (bases 1 to 1084)   |  |  |           |              |           |       |         |         |
| Genoscope.  |  |  |           |              |           |       |         |         |
| Direct Submission   |  |  |           |              |           |       |         |         |
| Submitted (07-SEP-2000)   | Genoscope - Centre National de Sequencage, |  |           |              |           |       |         |         |
| 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : |  |  |           |              |           |       |         |         |
| segrete@genoscope.cns.fr - Web : www.genoscope.cns.fr)              |  |  |           |              |           |       |         |         |
| This GSS is part of a random genomic sequencing program of thirteen |  |  |           |              |           |       |         |         |
| yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces     |  |  |           |              |           |       |         |         |
| exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,         |  |  |           |              |           |       |         |         |
| Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces |  |  |           |              |           |       |         |         |
| lactis var. lactis, Kluyveromyces hanseni var. hanseni, Pichia      |  |  |           |              |           |       |         |         |
| angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  |  |  |           |              |           |       |         |         |
| Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to |  |  |           |              |           |       |         |         |
| 5 kb were prepared and both extremities were sequenced. See         |  |  |           |              |           |       |         |         |
| keywords for description of this sequence and for the sequence of   |  |  |           |              |           |       |         |         |
| the other extremity of this insert.                                 |  |  |           |              |           |       |         |         |
| Location/Qualifiers   |  |  |           |              |           |       |         |         |
| 1..1084   |  |  |           |              |           |       |         |         |
| /organism="Saccharomyces exiguus"                                   |  |  |           |              |           |       |         |         |
| /strain="CBS 379"   |  |  |           |              |           |       |         |         |
| /db_xref="taxon:34358"  |  |  |           |              |           |       |         |         |
| /clone="AY0AA015D02"  |  |  |           |              |           |       |         |         |
| /clone_lib="AY0AA"  |  |  |           |              |           |       |         |         |
| /note="end : 17"  |  |  |           |              |           |       |         |         |
| /note="complement(<17..>979)"                                       |  |  |           |              |           |       |         |         |
| /note="similar to Saccharomyces cerevisiae ORF YMR304w [            |  |  |           |              |           |       |         |         |
| UBP5 : similarity to human ubiquitin-specific protease ]            |  |  |           |              |           |       |         |         |
| 1 putative frameshift(s)"   |  |  |           |              |           |       |         |         |
| /evidence="not_experimental"  |  |  |           |              |           |       |         |         |
| BASE COUNT  | 214 a                                      | 114 c  | 92 g      | 630 t        | 34 others |       |         |         |
| ORIGIN  |  |  |           |              |           |       |         |         |
| Query Match   |  |  |           |              |           |       |         |         |
| Best Local Similarity   | 45.0%;                                     | Score 72.8;  | DB 17;    | Length 1084; |           |       |         |         |
| Matches 191;  | Conservative 18;                           | Mismatches 215;  | Indels 0; | Gaps 0;      |           |       |         |         |
| OY  | 523  | TTTAAATCGTACGCTCATATTCANTAAATTCCTACTATTAATAAAGCTCTAAAGGC   | 582       |              |           |       |         |         |
| DB  | 118  | TTTATCTAATCTCTCCYCAATATATATTTTMMATTTTATTTATTTATTTCTTTTGT   | 177       |              |           |       |         |         |
| OY  | 583  | TGTTAATGATGATATTCGAATATTAACATATTAATTAATGCTATTAATTAATCTGCG  | 642       |              |           |       |         |         |
| DB  | 178  | TWTWATTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT | 237       |              |           |       |         |         |
| OY  | 643  | TATTTGATACAGCTTCATGCTCATATACATGTAATATATTTGAGTTAGACCTTAATCA | 702       |              |           |       |         |         |
| DB  | 238  | TT | 297       |              |           |       |         |         |
| OY  | 703  | AGGTAATTCGCTAATTAATATATATATCGAATTAATTAATTAATTAATTAATTAAT   | 762       |              |           |       |         |         |
| DB  | 298  | TT | 357       |              |           |       |         |         |
| OY  | 763  | TTTTATGTTTTCATTTTAAATGACGAGCTTGTGATTCATTAATTTTAAATGATGA    | 822       |              |           |       |         |         |
| DB  | 358  | TT | 417       |              |           |       |         |         |
| OY  | 823  | CAACATCTTTATGAGATTTAGATATTTGTAATGATGACGAGGCTTTGCGTATATTT   | 882       |              |           |       |         |         |
| DB  | 418  | TT | 477       |              |           |       |         |         |
| OY  | 883  | TTATATTAATCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA   | 942       |              |           |       |         |         |
| DB  | 478  | TT | 537       |              |           |       |         |         |
| OY  | 943  | TTTTT 946  |           |              |           |       |         |         |
| DB  | 538  | TTTTT 541  |           |              |           |       |         |         |

[illegible]



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2003, 20:02:01 ; Search time 63 seconds

(without alignments)  
979.286 Million cell updates/sec

Title: US-09-915-706a-2

Perfect score: 2343

Sequence: 1 MPLSKHQIEQLSKPLSDSI.....DRGSKQDQKQSSATSALSW 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: A.Geneseq\_101002.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 152.5 | 6.5         | 1503   | 19    | AAW48845    |
| 2          | 138   | 5.9         | 2008   | 18    | AAW22016    |
| 3          | 136.5 | 5.8         | 7201   | 22    | ABB71136    |
| 4          | 135   | 5.8         | 2013   | 22    | AAW67964    |
| 5          | 134.5 | 5.7         | 1179   | 22    | AAU37669    |
| 6          | 134.5 | 5.7         | 1179   | 22    | AAW01107    |
| 7          | 128   | 5.5         | 2017   | 22    | ABG06301    |
| 8          | 126.5 | 5.4         | 752    | 23    | AAU84329    |
| 9          | 126   | 5.4         | 610    | 19    | AAW68206    |
| 10         | 122   | 5.2         | 1370   | 22    | ABB68691    |

|    |       |     |      |    |          |
|----|-------|-----|------|----|----------|
| 11 | 122   | 5.2 | 1971 | 22 | ABG07508 |
| 12 | 120.5 | 5.1 | 1048 | 22 | ABW59245 |
| 13 | 120.5 | 5.1 | 1294 | 22 | ABB63502 |
| 14 | 120   | 5.1 | 1131 | 22 | ABG07281 |
| 15 | 119.5 | 5.1 | 1497 | 22 | ABG20153 |
| 16 | 119.5 | 5.1 | 1066 | 22 | AAW67418 |
| 17 | 119.5 | 5.1 | 1179 | 22 | AAU38015 |
| 18 | 119   | 5.1 | 1023 | 23 | AAW48189 |
| 19 | 118.5 | 5.1 | 806  | 21 | AAW77574 |
| 20 | 118   | 5.0 | 1392 | 20 | AAW06999 |
| 21 | 117.5 | 5.0 | 687  | 19 | AAW41586 |
| 22 | 117.5 | 5.0 | 868  | 22 | ABW59635 |
| 23 | 117.5 | 5.0 | 1427 | 12 | AAW10534 |
| 24 | 117   | 5.0 | 2779 | 22 | ABW62371 |
| 25 | 116   | 5.0 | 815  | 22 | ABW60653 |
| 26 | 116   | 5.0 | 861  | 22 | ABW20258 |
| 27 | 115.5 | 4.9 | 663  | 22 | AAW94117 |
| 28 | 115.5 | 4.9 | 1690 | 22 | ABW61144 |
| 29 | 115.5 | 4.9 | 1690 | 22 | ABW61173 |
| 30 | 115.5 | 4.9 | 2519 | 22 | ABG16636 |
| 31 | 114.5 | 4.9 | 955  | 15 | AAW57365 |
| 32 | 114.5 | 4.9 | 955  | 17 | AAW03691 |
| 33 | 114.5 | 4.9 | 1091 | 22 | ABW68898 |
| 34 | 114.5 | 4.9 | 2053 | 22 | ABW71118 |
| 35 | 114   | 4.9 | 952  | 22 | AAW79258 |
| 36 | 114   | 4.9 | 1939 | 23 | ABW77096 |
| 37 | 113.5 | 4.8 | 2542 | 22 | ABW71337 |
| 38 | 113.5 | 4.8 | 2816 | 22 | AAW68572 |
| 39 | 113   | 4.8 | 1411 | 17 | AAW02258 |
| 40 | 112.5 | 4.8 | 733  | 21 | AAW07449 |
| 41 | 112.5 | 4.8 | 1130 | 22 | ABG12663 |
| 42 | 112   | 4.8 | 881  | 22 | ABG05280 |
| 43 | 112   | 4.8 | 900  | 22 | AAW20178 |
| 44 | 112   | 4.8 | 900  | 22 | AAW20179 |
| 45 | 112   | 4.8 | 2013 | 22 | ABB62322 |

#### ALIGNMENTS

|          |   |
|----------|---|
| RESULT 1 |   |
| AAW48845 |   |
| ID       | AAW48845 standard; Protein; 1503 AA.                          |
| AC       | AAW48845;   |
| DT       | 12-OCT-1998 (first entry)                                     |
| DE       | Human receptor tyrosine kinase LMR2_h.                        |
| XX       |   |
| KW       | Receptor tyrosine kinase; LMR2_h; human; signal transduction; |
| KW       | cancer; neurodegenerative disorder.                           |
| XX       |   |
| OS       | Homo sapiens.   |
| XX       |   |
| FH       |   |
| FT       | Key   |
| FT       | Peptide   |
| FT       | Domain  |
| FT       | Location/Qualifiers   |
| FT       | 1..23   |
| FT       | /label= Sig_peptide   |
| FT       | 43..69  |
| FT       | /label= TMD   |
| FT       | /note= "Transmembrane domain"                                 |
| FT       | 409   |
| FT       | /label= O-phosphorylated                                      |
| FT       | 477   |
| FT       | /label= O-phosphorylated                                      |
| FT       | 552   |
| FT       | /label= O-phosphorylated                                      |
| FT       | 1032  |
| FT       | /label= O-phosphorylated                                      |
| FT       | 1100  |
| FT       | /label= O-phosphorylated                                      |
| FT       | 1105  |
| FT       | /label= O-phosphorylated                                      |

Novel human diago  
Drosophila melanog  
Drosophila melanog  
Novel human diago  
Novel human diago  
Novel human diago  
Amino acid sequenc  
Streptococcus pneu  
Listeria monocytog  
Human cytoskeletal  
Resilin protein seq  
Truncated resilin p  
Drosophila melanog  
Human 160kD mediat  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Human protein sequ  
Drosophila melanog  
Drosophila melanog  
Novel human diago  
K39 polypeptide of  
Leishmania chagasi  
Drosophila melanog  
Drosophila melanog  
Human protein seq  
Human alpha-myosin  
Drosophila melanog  
Human novel cytol  
Nucleolar/endosoma  
Amino acid sequenc  
Novel human diago  
Novel human diago  
S. cerevisiae mult  
D. cerevisiae mult  
Drosophila melanog

```
FT Modified-site 1365
FT /label= O-phosphorylated
FT Peptide 84...98
FT /note= "immunogen"
FT Peptide 684..698
FT /note= "immunogen"
FT Peptide 1098..1113
FT /note= "immunogen"
FT Peptide 1488..1503
FT /note= "immunogen"
FT Peptide 163..166
FT /note= "protein kinase conserved motif"
FT Peptide 281..283
FT /note= "protein kinase conserved motif"
FT Peptide 329..332
FT /note= "protein kinase conserved motif"
XX
XX
XX MO9822507-A2.
XX
XX 28-MAY-1998.
XX
XX 21-NOV-1997; 97WO-US2526.
XX
XX 22-NOV-1996; 96US-0031675.
XX
XX (SUGEN-) SUGEN INC.
XX
XX Johe KE, Plowman GD;
XX
XX WPI: 1998-312419/27.
XX
XX N-PSDB; AAV32452.
XX
XX
XX New isolated receptor tyrosine kinase genes - which are expressed in
XX neuronal tissues and tumour cells, useful as targets for
XX neurodegenerative disorders or cancers
XX
XX Example 1: Fig 7B; 87pp; English.
XX
XX This is the amino acid sequence of human LMR2-h, deduced from a
XX LMR2-h cDNA clone (see AAV32452). Novel rat, human and mouse LMR1,
XX LMR2 and LMR3 (see AAV48841-49) define a novel family of receptors
XX that are structurally related to receptor tyrosine kinases (RTKs).
XX They all share the distinct motifs that typically characterise this
XX class of enzymes, but possess extremely short extracellular domains
XX and have C-terminal tails of unprecedented length among RTKs.
XX Expression of LMR1 and LMR3 is highly restricted to neuronal
XX tissues with minimal expression in other adult or embryonic organs
XX or in human tumour cell lines. LMR2 expression is limited to adult
XX neuronal tissues, but is also very abundantly expressed in other
XX non-neuronal foetal tissues and in numerous tumour cell lines.
XX Based on restricted expression of all 3 LMRs to adult neuronal
XX tissues and the up regulation of LMR2 in a wide variety of tumour
XX cell lines, these proteins may be critical targets for
XX neurodegenerative disorders or cancer. Anti-LMR antibodies may be
XX used for detecting neurodegenerative diseases or cancer
XX
XX
XX Sequence 1503 AA:
SQ
Query Match 6.5%; Score 152.5; DB 19; Length 1503;
Best Local Similarity 21.4%; Pred. No. 0.0026;
Matches 122; Conservative 75; Mismatches 189; Indels 183; Gaps 25;
QY 3 LSKHQIQLSLSD--DSICGYLKLKSAFRLRNEFVVAQTALRKLSONPSADERD 59
D 370 LPRQLQLEQ--PYSDRWYEVQFCWLSPEK--RPAADVRLRLTLRLQSGORS--EVD 421
QY 60 ALQACLKKKIIISDLYEFSKTTREILISWVAAOF-----LDTTESAANSLEW 113
D 422 FEQD-----NNALPNTNSRDSNNAAFPILDFARDRLGREMEVILVTFETSGLSPEY 476
QY 114 LADISE-KHWD-----HNN-----PVLPEITLKSDDDKGERQDAKXKAFQQL 157
D 477 VWEAKHDFEDERSRGLDEGLSTSTIFYPVEFESSLSDPGKQ----- 522
```

```
QY 158 VGDEESSIIYAPVLQPLVGEVTFPD-----FQSAERKG---EISQLKMLTT 203
D 523 -DSGQDVPLRVP-----GVVPVFDANLTVSGSDYIQLEEKSGSLIEDYPPALLTY 574
QY 204 TVAQRFAIQFMENAKRCYQVLDRLSLVSTKCHSLGSGOSTNGCFKSLTLRENNLVH 263
D 575 -----DMDNPEKTGPQLSQTALRSVE---LEESSTDDEFPQSTDPKDSLLPG 620
QY 264 LSGIKLAPKAEKTVQEVQAEVSECELPMDTKHIERIPMASEQAQ---TVSQHLH 319
D 621 DLHTSGESEPNNIENDVKSE---DLPSHQ--KIFDLMELNGVQADFNPATLSSLD 674
QY 320 -----AGNISELGNINNMNRDLAFHLAREVSDYPROSE 352
D 675 NPKESVITGFPEKPKRKIFDSEPLCLSDNLMHODNFDPLN-----VQELSENF 723
QY 353 PHSPIFLLEKAIKWGISLPELLREKMSF-QNGDALSTIRNAGLNHLD----- 401
D 724 -----LFLQEKNLKLSLSKKEHINDLOTELKNAGFTFEMLETSCRSRLDELOFAENKP 778
QY 402 -----QVIL-----PEVSTPTVGEISPPQPAKPSVSDPSV 433
D 779 GLSLQENVSTKDDTDVMLTGDTLSTLSLOSSPEYQVPPSTFETETPRVPPDPLPQG 838
QY 434 EEHVSQTSPPVDTSQKQDQKQSSATSALS 462
D 839 E---TQPTCLDVIVPEDCIHQDISPDAVT 864
RESULT 2
AAW22016
ID AAW22016 standard; Protein; 2008 AA.
XX
XX AAW22016;
XX
XX 03-OCT-1997 (first entry)
XX
XX utrophin truncated polypeptide.
XX
XX utrophin; minigene; Duchenne muscular dystrophy; gene therapy.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Region 238..250
XX /note= "residues 238-250 (all encoded by codon NNN)
XX may comprise the sequence DKKSITMYTSL,
XX absolutely conserved in human, mouse and
XX rat utrophins"
XX
XX WO9722696-A1.
XX
XX 26-JUN-1997.
XX
XX 19-DEC-1996; 96WO-GB03156.
XX
XX 24-OCT-1996; 96GB-0022174.
XX
XX 19-DEC-1995; 95GB-0025962.
XX
XX 26-JUL-1996; 96GB-0015797.
XX
XX (MED-) MEDICAL RES COUNCIL.
XX
XX Davies KE, Tinsley JM;
XX
XX WPI: 1997-341687/31.
XX
XX N-PSDB; AAT74665.
XX
XX Nucleic acid encoding utrophin, truncated forms and related vectors
XX - also transformed mammalian cells, used for alleviating symptoms of
XX muscular dystrophy
XX
XX Claim 1; Fig 3; 78pp; English.
PS
```

XX A truncated utrophin polypeptide (AAW22016) having utrophin function  
CC includes the actin-binding domain and dystrophin protein complex  
CC (DPC) binding domain of full-length utrophin (see also AAW22017), but  
CC lacks the rod domain. It is obtd. by expression of a utrophin  
CC minigene (AAW74665) comprising approx. the first 2 kb and the last 4  
CC kb of the full-length utrophin coding sequence (see also AAW74666).  
CC Expression of the truncated utrophin significantly decreases the  
CC severity of the dystrophic muscle phenotype in an animal model.  
CC indicating usefulness in treatment of muscular dystrophy. The  
CC polypeptide can also be used to screen for substances that modulate  
CC utrophin binding to actin and/or the DPC.  
XX  
SQ Sequence 2008 AA:  
Query Match 5.9%; Score 138; DB 18; Length 2008;  
Best Local Similarity 21.2%; Pred. No. 0.063;  
Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19;  
QY 1 MPLSKHQLSEQLSDSDSICGVYLLKESAFRLRNFVNAQTALRKLSONSPADERDA 60  
DB 420 MELQKKQLQQLSSWLA-----LTERIQKMESEPLIGDULPS 455  
QY 61 LQEACLKMKLSDSLYEQSKTRDIELT-----SWFVAQFLIDPTTLESANSLWLA 115  
DB 456 LQK-LLEHKSILQNDLEAEQKVNSTLHMVYIVDENSEGSAITALLDQILQKGRMTAVC 514  
QY 116 DLSEKHHDLNPLVPVETLKSDDKGRERQADAKVAFQVLDSESSITLYAPVQLP 175  
DB 515 RWTERNNRL-----DEISILWQELLEEQ 538  
QY 176 LVGEVTFEFDQSAERKEISQLKSMLTITVAOERFAIOFKMENAKRCVTOIDRLSAL-- 232  
DB 539 CLLEAWLTERKEEALNKTQVSFKDQKELSVSRRLAILKEEMKRR--QITDQLEISQGD 596  
QY 233 -----VSTKCHSGSOSTNGFSAKSLITRYENALVHLGKILAPKAEAKTVEGEVA 283  
DB 597 VGQLLSNPKASKKNSSEELTORW--DSLQVRLEDS-----SNQVTOAVA 640  
QY 284 ESSVEGELPSHMDTKHIERIPMASEOQ-----TVSOH-LHAGNLSELGNLNNMRDL 336  
DB 641 KLGMSQIPQKDLLETIVHREKGVKKRKQELPPLTKAEHAKMKRSTTELG----- 691  
QY 337 AFHLLREVSDFROSEPH-----SPISFILLEKAIKMGYSLPELLREMMSEONGDA 387  
DB 692 --EMIQRLRLTOEMEVHAEKRLKWLNRTELEMLSDKS-----LSLPE--RDKISE--S 738  
QY 388 LSTFENAGLNHLDOVLLPEVSTPTVGIESPQTFOAKPVSQDPSVEBHVSQTSVPDTQS 447  
DB 739 LRTV-----NMTWNKICREV--PTTLKECIOEPS--SVSQTR-IAHNPVQKVLVSS 786  
QY 448 KODKPOSSATSALS 462  
DB 787 ASDIPVOSHRTSEIS 801  
RESULT 3  
ID ABB71136 standard; protein; 7201 AA.  
XX ABB71136;  
XX 26-MAR-2002 (first entry)  
DE Drosophila melanogaster polypeptide SEQ ID NO 40200.  
KW Drosophila; developmental biology; cell signalling; insecticide;  
OS pharmaceutical.  
XX Drosophila melanogaster.  
XX W0200171042-A2.  
XX

PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001MO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL15239.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 40200; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 7201 AA:  
Query Match 5.8%; Score 136.5; DB 22; Length 7201;  
Best Local Similarity 22.0%; Pred. No. 0.54;  
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;  
QY 30 SAPRRLNE-FNVQTA-----LRKLS-----QNSAD---ERD---ALQEA---CLKMK 70  
DB 1263 SEFDQKNDIFSMLOTAELELRSLTPIOTDPKKNVSQDKSKSRDLNVQLOQASHOLRLKH 1322  
QY 71 ILSDLSQFSKTRDIELISWFAAOFLDPTTLESANSLWLADESEKHMW----- 124  
DB 1323 ALKSELRLAPARDKRP-LLESEVEVEKMFNTMEHKKDRGYLEDYSAK-WNNYKTRLA 1380  
QY 125 -----LNPVLP--VETLSDDDKGRQADAKVAFQVLD--SESSITLYAPVQLP 175  
DB 1381 ELQEMANKVAPKNEIALQSEDLTPREHV--VKQAFKRITLGRMKOLDLLADASELAP 1437  
QY 176 LVGEVTFEFDQSAERKEISQLKSMLT----- 203  
DB 1438 KEGNIA---EAKRLKEGILTQLEVLSAINNNDHQAQAVQEDLVNMQQFOAGLOQIKPA 1493  
QY 204 -----TVAOERFAIOFKMENAKRCVTOIDRLSALVSTKCHSGSQ 243  
DB 1494 VQSEVAVNNVWVSKRPILEEVAVMAQNAQOPEFO-----COBQDLKLG-ISTNISHMKLCK 1548  
QY 244 STNFGAKSLTRYEN--ALVHLSGILAPKAE-----KTVEGEVA--ESSVEGE-- 291  
DB 1549 -TN---APDELDAHNSHWTAVHENAQASAKLEKLVANMKSFDAKALIDBNVQGOBQM 1604  
QY 292 --LPSHMDTKHIERI-----PMASEQAQV-----SOHLHAGNLSEL-- 326  
DB 1605 SRRAVAVLNTPHIDLKELKLVKLSFNNEISQQAQKLTTCOMADQISLHLAPEGAAALKD 1664  
QY 327 -----GNLNNMRDLAFHLLREVSDF--ROSEPHSPISFILLEKAIKMGYSLPELLRE 378  
DB 1665 RVNQMKKGLKRLSATRGH--INEVSDAIIISROFNALVNF-----SNW-----MEQLRN 1713  
QY 379 MMS---EQNGDALST---INAGLNHLDO-----VLLPEVSTPTVGIESPQTFOAKPSV 427  
DB 1714 QVTOVEINPERVETSLHVIHALQEHADKKPSPAIYDEVKQALAG---ATPEESNAL 1769

```

OY 428 SDP-----RSVGEHSHQTSFV-----DPOS-----KODQKPOSSATSL 461
Db 1770 NDATAYALVNTIONLETNMLQKKALEKWTELLGKNDTESHUNLYKHOLDKPEGPAEEL 1829
OY 462 S 462
Db 1830 S 1830

RESULT 4
AAB67964
ID AAB67964 standard; Protein; 2013 AA.
XX
XX AAB67964;
AC
XX
XX 29-JUN-2001 (first entry)
DT
XX
XX Amino acid sequence of utrophin B isoform minigene.
DE
XX
XX utrophin; promoter; chromosome 6q24; exon 1b; dystrophin;
KM muscle-specific transcription; muscular dystrophy.
XX
XX Unidentified.
OS
XX
XX W0200125461-A1.
PN
XX
XX 12-APR-2001.
PD
XX
XX 04-OCT-2000; 2000WO-GB03800.
PF
XX
XX 04-OCT-1999; 99GB-0023423.
PR
XX
XX (ISIS-) ISIS INNOVATION LTD.
PA
XX
XX Burton E, Tinsley J, Davies K;
PI
XX
XX WPI; 2001-273582/28.
DR
XX
XX NP-PSDB; AAF84673.
XX
XX
XX Novel nucleic acid comprising promoter for mouse, human utrophin genes,
PT for promoting tissue-specific transcription of linked sequences and in
PR screening for substances able to modulate utrophin promoter activity -
XX
XX Disclosure; Fig 9; 76pp; English.
PS
XX
XX The present sequence is encoded by an utrophin B isoform minigene.
CC The specification describes human and murine utrophin alternative
CC promoters. utrophin is a 395 kDa protein encoded by a gene located on
CC chromosome 6q24. The alternative promoter is highly regulated, and has
CC little similarity to the synaptically expressed promoter. The
CC alternative promoter drives transcription of a widely expressed unique
CC first exon that splice into a common full length mRNA at exon 3. This
CC unique exon (called exon 1b) encodes a novel 31 peptide which may be
CC involved in binding to the muscle membrane. utrophin alternative
CC promoters are useful for promoting tissue-specific, preferably
CC muscle-specific transcription of an operably linked sequence of
CC nucleotides. utrophin exon 1b polynucleotides and polypeptides are
CC useful in the manufacture of medicament for treating a dystrophin
CC phenotype in a mammal. Up-regulation of utrophin expression may
CC compensate for dystrophin loss in muscular dystrophy patients.
CC
XX
XX
SQ Sequence 2013 AA;
OY Query Match 5.8%; Score 135; DB 22; Length 2013;
Best Local Similarity 21.2%; Pred. No. 0.11.
Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19
OY 1 MFLSHQIEQLSEKPLSDSDSICGVYLLKLEKSARPLRNFENFVATLRKLSQNPSPADERDA 60
Db 425 MFLQKQKQLQDQLSSWLA-----NVAATL-----LTERKQKNEISPLGDDLP 460
OY 61 LOEACILMKRKILSDSLYEFQSKTYTDIELI-----SWFVAAGFLDPTLTLESANSLIEWIA 115

```

|          |  |  |      |  |  |
|----------|--|--|------|--|--|
| Db       | 461  | LQK-LDGHKRLQNDLDEAQYKVNSLTHMYIYDENGSESAFTALLEQDLKGERWTAVC    | 51.9 |  |  |
| Oy       | 116  | DLSEKHMDHLNPLVPLEVTFLKSDDDKKEREQADAKVKAFFOLYGDSESESLTAPVLQLP | 17.5 |  |  |
| Db       | 520  | RWTEERRNRL-----OESILWOELLEQ                                  | 54.3 |  |  |
| Oy       | 176  | LVGEYTFEPDQSAERKGEISQLKSMILTTVAQERFAIOFKMENKRCVTOLODRLSAL--- | 23.2 |  |  |
| Db       | 544  | CLEEMWLTEKEEALDKVQTSNFKQDKELSVYRKLALIKEDEMKR--QTLDDLSLEIGD   | 60.1 |  |  |
| Oy       | 233  | -----VSTKCHSLGSGSTNGEFAKSLTTRVENALVHLGSIKLAPKAERAKTEQEVA     | 26.3 |  |  |
| Db       | 602  | VGOLLNPKASKKNNMSDESLTORW--DSLWORLEDS-----SNOQTQAVA           | 64.5 |  |  |
| Oy       | 284  | ESSVSEGLPSPHMTKHTERTPMASEQAQ-----TVSQH-LHNGNLSLGNLNNMRDL     | 33.6 |  |  |
| Db       | 646  | KLGMSQIPQKDLLELVHAREGQWKKRPQELPPLTGAENHAKRSTTELG-----        | 69.6 |  |  |
| Oy       | 337  | AFHLLREVSDFRQSEPH-----SPISFLEKAIIRMGYLSLPRLREMMSEONGDA       | 38.7 |  |  |
| Db       | 697  | ---ENQGLRDLTQEMEVHAEKLMNLNRELEMLSDKS-----ISLPE--RDKISE---S   | 74.3 |  |  |
| Oy       | 368  | LSTIFNAAGLNLHLDVLLPEVSTPTVGIESQTQOAKPVSDDPSVDEHNSQTSRPVDTQS  | 44.7 |  |  |
| Db       | 744  | LRTV-----NMTWNKLCREV--PTTLKECIOEPS--SVSQTR-IAHPNVQKVVLVSS    | 79.1 |  |  |
| Oy       | 448  | KODQKPOSSATSALS 462  |      |  |  |
| Db       | 792  | ASDIPVQSHRTSEIS 806  |      |  |  |
| RESULT 5 |  |  |      |  |  |
| AAU37669 |  |  |      |  |  |
| ID       | AAU37669   | standard; Protein: 1179                                      | AA.  |  |  |
| XX       | AAU37669;  |  |      |  |  |
| XX       | AC   |  |      |  |  |
| XX       | DT   |  |      |  |  |
| XX       | 14-FEB-2002  | (first entry)  |      |  |  |
| DE       | Streptococcus pneumoniae cellular proliferation protein #98.     |  |      |  |  |
| XX       |  |  |      |  |  |
| KW       | Antisense: prokaryotic cellular proliferation protein;           |  |      |  |  |
| XX       | antibiotic; antibacterial; drug design.                          |  |      |  |  |
| OS       | Streptococcus pneumoniae.  |  |      |  |  |
| XX       |  |  |      |  |  |
| PN       | WO200170955-A2.  |  |      |  |  |
| XX       |  |  |      |  |  |
| PD       | 27-SEP-2001.   |  |      |  |  |
| XX       |  |  |      |  |  |
| FE       | 21-MAR-2001; 2001WO-US09180.                                     |  |      |  |  |
| XX       |  |  |      |  |  |
| PR       | 21-MAR-2000; 2000US-191078P.                                     |  |      |  |  |
| XX       |  |  |      |  |  |
| PR       | 23-MAY-2000; 2000US-206848P.                                     |  |      |  |  |
| XX       |  |  |      |  |  |
| PR       | 26-MAY-2000; 2000US-207727P.                                     |  |      |  |  |
| XX       |  |  |      |  |  |
| PR       | 23-OCT-2000; 2000US-242578P.                                     |  |      |  |  |
| XX       |  |  |      |  |  |
| PR       | 27-NOV-2000; 2000US-253625P.                                     |  |      |  |  |
| XX       |  |  |      |  |  |
| PR       | 22-DEC-2000; 2000US-257931P.                                     |  |      |  |  |
| XX       |  |  |      |  |  |
| PR       | 16-FEB-2001; 2001US-269308P.                                     |  |      |  |  |
| XX       |  |  |      |  |  |
| PA       | (ELIT-) ELITRA PHARM INC.  |  |      |  |  |
| XX       |  |  |      |  |  |
| PI       | Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ, |  |      |  |  |
| XX       |  |  |      |  |  |
| PI       | Yamanoto RT, Xu HH;  |  |      |  |  |
| XX       |  |  |      |  |  |
| XX       | WPI; 2001:611495/70.   |  |      |  |  |
| DR       | N-PSDB; AAS55528.  |  |      |  |  |
| XX       |  |  |      |  |  |
| XX       |  |  |      |  |  |
| PT       | New polynucleotides for the identification and development of    |  |      |  |  |
| XX       | antibiotics, comprise sequences of antisense nucleic acids -     |  |      |  |  |
| XX       |  |  |      |  |  |
| PS       | Example 3;Seq ID No 13262; 511pp; English.                       |  |      |  |  |

XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
SQ Sequence 1179 AA:

Query Match 5.7%; Score 134.5; DB 22; Length 1179;  
Best Local Similarity 21.6%; Pred. No. 0.057;  
Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

QY 5 KHOIEOLSKPLSDSICGVYLLKLEKSAFRLPRNFENNAQTALRLKLSQNSPADERDALQEA 64  
DB 674 KPELEQLOKETAIDE---ASIGSEEAALKTLQDDMAALTERLEAIK---SOGEGARIOEQ 727  
QY 65 CLNWKILSLDYEQFSKTTIDILISMVFAAQFLDITLESANSLEMLADSEKHMDH 124  
DB 728 GLS-----LAYOOTSQOYEELETL-WKLOEE-EIDRLSG-----DWQAD-KEKCOES 772  
QY 125 L-----NPLVPELTAKSDDDKGRERQADAKVKAFFOLVGSSESSILYAPVLOPL 176  
DB 773 LATIASKONLEAEIEIKSKNAIOERYQ-----NLQEEVAQARILKTEL 818  
QY 177 VGEVTFDFQSAERKGEISQLSKMLTTVAQERF-AIOFKMENAKRCVQDLRLSALVS 234  
DB 819 QGQR-YEVADIERLG-----KELDNLNIEOEIORMLOEKVDNLEKYDE-----LLS 866  
QY 235 TKCHLSGSOSTNF--GFAKSL--LTRVENALVHLSG-----IKLAPRAEAKTV 278  
DB 867 QQAESKTQKTNLQOGLIRKOFELDDIEGQLDILASHLDQARQONEEMIRKQTRAERK-- 924  
QY 279 EQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNIS--ELGNLNMNRDL 336  
DB 925 KEKVSERL-----RHQNLQDQOQISYTALEKAHLEMLNLAEQVEDDEKAIKRL 977  
QY 337 AFHLRLREVSDFRQSEPHSPISFLLEKAIKRGVLSPELLREMSSEONGDA--LSTTFN 393  
DB 978 GPNVLEAIDQY---EEVHNRLDPL--NSQRDILISAKNLLLETITENMDEVKEREKSTFE 1032  
QY 394 AA-----GLNHLDOVLLEPVESTPTVGIESPOTPOAK 424  
DB 1033 AIRESEKVTFRKQMGGOAD-LITLEDGLTAGVIEISVQPPGK 1074

RESULT 6  
AAM01107  
ID AAM01107 standard; Protein; 1179 AA.  
XX

AC AAM01107;  
XX  
XX 02-OCT-2001 (first entry)  
XX  
XX CFE 110 protein sequence.  
DE  
KM Antibacterial; vaccine; gene therapy; bacterial cell wall viability;  
KM CFE; Conserved Essential Gene; bacterial infection;  
KW antisense therapy; antibiotic resistance.

XX  
OS Streptococcus pneumoniae.  
XX  
XX WO200149721-A2.  
XX  
XX 12-JUL-2001.  
XX  
XX 29-DEC-2000; 2000WO-US35604.  
XX  
XX 30-DEC-1999; 99US-0174089.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Brucoleri RE;  
XX  
XX Thanassi JA;  
XX  
XX WPI; 2001-496721/54.  
XX  
XX N-PSDB; AAH90806.  
XX  
XX Nucleic acids encoding conserved essential genes involved in bacterial  
XX replication which are potential targets for the treatment of antibiotic  
XX resistant bacterial infections -  
XX  
XX  
XX Claim 27; Pages 365-369; 380pp; English.

The present invention relates to nucleic acids (AAH90701-AAH90918)  
CC encoding polypeptides (AAM01002-AAM01114), which are essential for the  
CC viability of a bacterial cell wall. The acronym CFE stands for "CEG For  
CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic  
CC acids are useful for detecting the presence of proteins essential for the  
CC viability of a bacterial cell wall in samples such as cells, tissues,  
CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,  
CC and for detecting corresponding target nucleic acid molecules with  
CC complementary sequences. The nucleic acids are also useful for  
CC determining whether a genomic nucleotide sequence of interest is  
CC essential for viability of a bacterial cell or whether it resides within  
CC an operon, by integrating an exogenous nucleotide sequence comprising a  
CC portion of an open reading frame of the genomic sequence of interest  
CC (comprising 200-500 base pairs) into the genomic sequence of interest  
CC which confers a selectable phenotype to the cell, and determining cell  
CC viability with a selection agent such as chloramphenicol. The nucleic  
CC acids and proteins are also useful as vaccines and for treating bacterial  
CC infections with gene therapy and antisense therapy. The nucleic acids  
CC also enable identification of targets suitable for the treatment of  
CC antibiotic resistant bacterial infections.

SQ Sequence 1179 AA:  
Query Match 5.7%; Score 134.5; DB 22; Length 1179;  
Best Local Similarity 21.4%; Pred. No. 0.057;  
Matches 99; Conservative 80; Mismatches 180; Indels 103; Gaps 23;

QY 5 KHOIEOLSKPLSDSICGVYLLKLEKSAFRLPRNFENNAQTALRLKLSQNSPADERDALQEA 64  
DB 674 KPELEQLOKETAIDE---ASIGSEEAALKTLQDDMAALTERLEAIK---SOGEGARIOEQ 727  
QY 65 CLNWKILSLDYEQFSKTTIDILISMVFAAQFLDITLESANSLEMLAD-----L 117  
DB 728 GLS-----LAYOOTSQOYEELETL-WKLOEE-EIDRLSG-----DWQADKEKCOESL 773  
QY 118 SEKHMDHLPVLPETLKSDDDKGRERQADAKVKAFFOLVGSSESSILYAPVLOPLV 177  
DB 774 ATIASDKONLEAEIEIKSKNAIOERYQ-----NLQEEVAQARILKTELQ 819  
QY 178 GEVTFDFQSAERKGEISQLSKMLTTVAQERF-AIOFKMENAKRCVQDLRLSALVST 235  
DB 820 QGQR-YEVADIERLG-----KELDNLNIEOEIORMLOEKVDNLEKYDE-----LLSQ 867  
QY 236 KCHLSGSOSTNF--GFAKSL--LTRVENALVHLSG-----IKLAPRAEAKTV 279  
DB 868 QAESKTQKTNLQOGLIRKOFELDDIEGQLDILASHLDQARQONEEMIRKQTRAERK--K 925  
QY 280 QEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNIS--ELGNLNMNRDLA 337

Db 926 EKVSRL-----RHLONLTDQOISYTEALEKAHELENNINLAEOQVODLEKAIKRLSG 978  
Oy 338 FHLLREVSDFRQSEPHSPISFLLEKATRWGLVSLPELLREMSQONDA---LSTINA 394  
Db 979 FVNLTAIDQY---EEVNRRLDPL--NSQRDILSKNLLLELTTEMNDEKEREKSTEEA 1033  
Oy 395 A-----GLNHDQVLLPEVSTPTVGIESPQTPQAK 424  
Db 1034 IRESFKVTFKQMGCGGQAD-LILTFEGDILLTAGVELSVQPPCK 1074

RESULT 7  
ABG06301  
ID ABG06301 standard; Protein: 2017 AA.  
XX  
AC ABG06301;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #6292.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
PA (HXSE-) HXSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
DR N-PSDB: AAS70488.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 36660; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG0010-ABG30377 represent novel human  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIGO  
XX at ftp.wigo.int/pub/published\_pcc\_sequences.  
XX  
XX Sequence 2017 AA;  
XX  
XX

Query Match 5.5%; Score 128; DB 22; Length 2017;  
Best Local Similarity 21.0%; Pred. No. 0.43;  
Matches 126; Conservative 77; Mismatches 191; Indels 206; Gaps 26;  
Oy 4 SKHOIEQLSKPLSDSISGCVYIKL-----EKS-----AFRLRNEFNVAQ 43  
Db 465 AENEIMRLSSINQNSLAEADNKLKMRLEVELEKESLISOKELOMSLILNNEYEVIK 524  
Oy 44 -TALRKLS-----ONPSADERDALOEA 64  
Db 525 STATPDISLSELDLRLNLAKKEOLNQSISEKETLIAETLELDROQAEFTKIMILKA 584  
Oy 65 CLNKWKILSDSLYEQFSKTPRDIEI-----ISMFYAAQFLDPTLTESANSU- 111  
Db 585 QLSKOQNEGDSIISKLODLNDEKKRVHQLEDKMDIKELDQGVFVLLIOSEVALNDLHL 644  
Oy 112 --EWLADLSEKHWHDLPVLPVETLKSDDDKGKEREQADAKVKAFFOLVGDSESSIIYA 169  
Db 645 TKOKLEDKVENLVDQLN-----KQSENVSTIOKEWLEKEHRIQ-----NEEELSRIRN 693  
Oy 170 PVLQ-LPLVGEVTPFDPSAEKGEISOLKSMLT-----TTVAQREFAIQEMENAK---- 220  
Db 694 ELMOSLNODSNSNEKDYTLKEREAEVRNLKONLSELOLNENLKKVAFDVMENKLVLA 753  
Oy 221 -----RCVTQDLRLS-----ALVSTKCHSLGSQSTNFGAKSLT-----TRVENALVH 263  
Db 754 CEDVHQHLEBCLAGNNOISLEKNTIVETLKMKEGIEALELQWAKKRILLEANKYEKTEE 813  
Oy 264 LSG-----IKLAPKAEAKTYEQEVAESSVSEGBGLPSMDTKHIERIPMA 307  
Db 814 LSNARNLNTSALOIEHEHLIKLNQK-----DMEIAELKKN-----IEQMDHRETKDVL 864  
Oy 308 SEQAGTVSCHLHAGNLSELGMINNMNRRDLAFHLLREVSDYRQSPHSPISFLEK---A 364  
Db 865 SSSLEEQKQ-----LTQLN-----KKEIFTEKLEKSSKIQEE-----LDKTSQA 905  
Oy 365 IRWGYLSPELLRE-----MMSQNGDALSTIFNAGLNHLQVLT--LPEVSTPT 412  
Db 906 LRKN-----ETLRQTEIEKDRSLSGMKEN-----NHLQBELRLREBQSR 947  
Oy 413 VGIESPQT-----PQAKPSVSPRSVEEHV-----SOTSFVDTQSKODQRPQ 454  
Db 948 APVADPKTIDSVTELAISEVQSLNITKEHLEEEIKHHOKIIEDQNSKQLQSLQEOKE 1007

RESULT 8  
AAU84329  
ID AAU84329 standard; Protein: 752 AA.  
XX  
XX AAU84329;  
XX  
XX 08-MAY-2002 (first entry)  
XX  
XX Protein DCC1 differentially expressed in breast cancer tissue.  
XX  
XX Human: diagnosis of breast cancer; endometrial cancer; breast tumour;  
KW MAI: mitotic activity index; cytostatic.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200210436-A2.  
XX  
XX 07-FEB-2002.  
XX  
XX 27-JUL-2001; 2001WO-US23642.  
XX  
XX 28-JUL-2000; 2000US-222093P.  
XX  
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
PA (BAK/) BAK J.  
XX  
XX Baak J, Mutter GL;  
PI



```

OY      257 VKNALVYLSGKILPAEKKTEVGEAEASSVSNGELPSHMDTHIERITPMASDAQVTSQ 316
        || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      245 VEEGLEIUSGHILDKADL-TKDIKALESNVEGLL--DISGLKADKADLTDKIDALES 301
OY      317 HLHAGNLSLGNLNNNRDLA-----FHLREVSDFYFRQSEPHSPISFLLEKAIKRWG 368
        :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      302 NVEEGHLDLSGRLLDDKADIAQNGTDIODLAAVNEILOYAKQTEAIDA--LNKASSEN 359
OY      369 YLSLEPL-----LREMSHQNGDALSTIRNAGLN 398
        :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      360 TQNIEDLAAYNELQDAYAKQTEAIDALNKASSEN 394

RESULT 10
ABB68691
ID      ABB68691 standard; Protein; 1370 AA.
XX
XX      ABB68691;
AC
XX      26-MAR-2002 (first entry)
DT
XX      Drosophila melanogaster polypeptide SEQ ID NO 32865.
DE
XX      Drosophila; developmental biology; cell signalling; insecticide;
KW      pharmaceutical.
XX
XX      Drosophila melanogaster.
OS
XX      WO200171042-A2.
PN
XX      27-SEP-2001.
PD
XX
XX      23-MAR-2001; 2001WO-US09231.
PF
XX
XX      23-MAR-2000; 2000US-191637P.
PR
XX      11-JUL-2000; 2000US-0614150.
XX
XX      (PERE ) PE CORP NY.
PA
XX
XX      Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX      WPI; 2001-656860/75.
DR
XX      N-PSDB; ABL12794.
XX
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
Pr      genes from Drosophila and for elucidating cell signalling and cell-cell
Pr      interactions -
XX
XX      Disclosure; SEQ ID NO 32865; 21np + Sequence Listing; English.
PS
XX
XX      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
CC      sequences (AB57737-AB872072).
CC      (AB57737-AB872072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 1370 AA;
SQ

```

|                       |                  |                 |             |              |
|-----------------------|------------------|-----------------|-------------|--------------|
| Query Match           | 5.28;            | Score 122;      | DB 22;      | length 1370; |
| Best Local Similarity | 21.68;           | Pred. No. 0.78; |             |              |
| Matches 130;          | Conservative 80; | Mismatches 176; | Indels 216; | Gaps 29;     |

```

07 3 LSKHOI-----EQLSKLSLSDSDSICGYLLEKKAEPRLRHEEVAQTLRAKLSQNSAD 56
    ||| ||| :||| : : : ||| ||| |||
Db 440 LSIHSMMAVEEQLAS--SNDSSMSASPQPSSEA--MRLAFSKAKQLRDLQMQVOSAS 494
    ||| ||| ||| ||| ||| ||| ||| |||
07 57 -----ERDLAQEACL--NKWKLSLSLSLYEQFS-- 81

```

|    |     |   |      |
|----|-----|---|------|
| Db | 495 | AVAQVAALIGELHNSMALPGFTTEQOLVGLDRQTCWMTLCAHGBAYALLNQLELPEFSQDM     | 554  |
| QY | 82  | -KTRDLELISWEEVAQAOFILDTTIESAANSLEWLADI,SEKHMHDLNVLEPELTKSDD-      | 1399 |
| Db | 555 | PEPSODVEVYN-----IFKLDSDDYVSVAFESJOTRLDKFPSTISLL--ENWVKDESL        | 6070 |
| QY | 140 | -----KKEKREQADAKYKAFEPVOLVGDSESSILYAPVLQ                          | 1738 |
| Db | 608 | ISIALHLSRQASIMQRFARTIKVMEPRKAYEELDAQYKAFOL-----LIS                | 6544 |
| QY | 174 | LPVGEVTFEFDQSAERKE-----ISOLKSMLTITVAQERFALQFMENAKRCVQ             | 2255 |
| Db | 655 | LPA-----QVANRLGRRLPETAFAVSQKLLRLQWLKSLHFVLOQ---DDNREYFD           | 7020 |
| QY | 226 | LDRLSAIYSTCHSLSGOSTNFGAKSILTFVEN-----ALVHLSGIKLPAKAKT             | 2777 |
| Db | 703 | LEPYSMWLISQAINLITYDVST---LESILRYLAKDYAAPRGKVVHTILKELDPAACLKT      | 7586 |
| QY | 278 | VEOEVAESSVEG-----ELPSIHDTYH-----IERIPMASEQAOVYSQHL                | 318  |
| Db | 759 | -----AQALASAGLNYLYLIGATLEP---HMKHCLLOKLPICQRTPVDMKCOLITLASYL      | 8110 |
| QY | 319 | HAGNLSSELGN,NMNRDLAFHLL-----REVSDFYRQSEPHSPISFLLEKAIKW-----G      | 3686 |
| Db | 811 | NAVPAQOVLIN-----OLLGWSKRISLOKLGSEOHIALSKLLVYLAKCJGNLYM            | 8633 |
| QY | 369 | YLSPELLEKEMSE--ONGDAL-----STEFNAGNLNHDVOLLPEVST-----              | 4110 |
| Db | 864 | DLINORQLHDGLSNHLQSPDLLOKHVGKMYELLFNF-----IALPRAKEDDLRFE           | 9155 |
| QY | 411 | -----PTVG--IESPQTPQ--AKRVSVDPRSVBEHVSQTSVPDTQSKODQ                | 4511 |
| Db | 916 | YDSFQDTFHHMHI,FEFBEFDLQCFESPSPKTKLEDKPCEDQLKOLELHLSDFMST--TEQKEPH | 9749 |
| QY | 452 | KP  | 453  |
| Db | 975 | KP  | 976  |

|    |   |
|----|---|
| xx | RESULT 11   |
| ID | ABG07508 standard; Protein; 1971 AA.                              |
| xx | ABG07508  |
| AC | ABG07508;   |
| xx |   |
| DT | 13-FEB-2002 (first entry)   |
| xx |   |
| DE | Novel human diagnostic protein #7499.                             |
| xx |   |
| KW | Human; chromosome mapping; gene mapping; forensic;                |
| RW | food supplement; medical imaging; diagnosis; genetic disorder.    |
| xx |   |
| OS | Homo sapiens.   |
| xx |   |
| PN | WO200175067-A2.   |
| xx |   |
| PD | 11-OCT-2001.  |
| xx |   |
| PF | 30-MAR-2001; 2001WO-US08631.                                      |
| xx |   |
| PR | 31-MAR-2000; 2000US-0540217.                                      |
| PR | 23-AUG-2000; 2000US-0649167.                                      |
| xx |   |
| PA | (HYSE-) HYSEQ INC.  |
| xx |   |
| PI | Dzmanac RT, Liu C, Tang YT;                                       |
| xx |   |
| DR | WPI: 2001-639362/73.  |
| xx |   |
| DR | N-PSDB; AAS71695.   |
| xx |   |
| PT | New isolated polynucleotide and encoded polypeptides, useful in   |
| xx | diagnostics, forensics, gene mapping, identification of mutations |

PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 20: SEQ ID No 37867; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 1971 AA:  
Query Match 5.2%; Score 122; DB 22; Length 1971;  
Best Local Similarity 19.5%; Pred. No. 1.3;  
Matches 103; Conservative 95; Mismatches 210; Indels 120; Gaps 25;  
14 PLSDSDICGVYLR-LKSAFRLPLNERNVAOTLRKL----- 49  
DB 239 PATGSSVMGTQMMNIGOSAPDVCSNE-DLPEVELVSLLEBQLROYLRKYDTLLEYNQDW 297  
QY 50 SQNPASDER--DALQACLN--KWKILSDLYEFSKTRDIELLSFWAAQFLDPTTL 104  
DB 298 TQSPHQHQSADLSPLABETFRMYLIGDRVEQMKTNDMDVYHLLAER---DRDL 354  
QY 105 ESNANSLLEMLAD-----LSEH-----WDHNPVLEVEYTLKSD-----DKKERE 145  
DB 355 ELAARIGQALLKRNNHVLSEONESLEBQLGQAFDVQNO-LQHEICKKDELRLIVASIEES 413  
QY 146 QADAKYKAFQVLVGDSESSILYAPVQLPLVGE-VTFEFGQSERGELISQKSMLT 204  
DB 414 ETBSSCSTPLRF-----NESLSISGLQLQLEMLQKLELEENNALRSKCHITETVTY 469  
QY 205 VADERFAIOFKMENAKRCYTQDLRLSALVSTK-----CHSLGSSSTNFGFA 250  
DB 470 EEKEQQLVSCVKELRETNNQMSMPELSEKSDXEDTKSPFLRHLXTFTPHKIGIT 529  
QY 251 KSLITRENALVHLSCGLKPAKAEKVEQVVAESSYSEBELPSHMDTKH---IERIPMA 307  
DB 530 KEOLT--GKYLPH---LPLSIENNLNQNFSISITK-EMLNRLSESEKTKLEQLHIM 581  
QY 308 SEQAOT--VSQHLHAGNLSELGNLNNMNRDLAF-HL---LREVSDFROSEPHSPISFL 361  
DB 582 QEOQKSLIDIGNQNVSEETVYINGNOQIDKVFSHIGADLLTGSDSNKED----- 632  
QY 362 EKAIRMGYLSLPELRLKEMSEQNGDALSTIFNAGLNHLDOVLLPEVSTFTVGIESPO-- 419  
DB 633 -----GALNVPRPAGAKPTQQRPTD-----MSALNNLFGPOKPKYSMMQLSQOKRNM 679  
QY 420 -----TQQAPSV--SDPRSYEHNVSOTSPVDTSQKQOKQSSATSL 461  
DB 680 LNFQVPQVSPATGSSVMGTQMMNIGOSAPDVCSNED-LPEVELVSL 726  
RESULT 12  
ABBS9245  
ID ABBS9245 standard; Protein; 1048 AA.  
XX  
XX  
AC ABBS9245;  
XX  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 4527.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
OS Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001MO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
DR N-PSDB; ABL03348.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure: SEQ ID NO 4527; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 1048 AA:  
Query Match 5.1%; Score 120.5; DB 22; Length 1048;  
Best Local Similarity 19.6%; Pred. No. 0.7;  
Matches 95; Conservative 78; Mismatches 162; Indels 149; Gaps 21;  
5 KHQEQLSKPLSDSIC-----GYLKLKSAFRLPLNERNVAOTLRKLSQNPASDER- 58  
DB 373 KEYTEELDK-LKRDLMARDKNGIYLA-----EETYGKITLKSQNELNEKM 420  
QY 59 ----DALQACLNKWKILSD--SL-----YEQSKTRDIELLSFWAAQFLDPTLES 106  
DB 421 LLKALKDELQNKKEITSEVSMISGLTLLTKKVLTKKRRYKKEKELVASHMKTEQVLT 480  
QY 107 AANSLLEMLADLSEKHWDLNPLVLP---VETLKSDDDKGERQADAKYKAFQVLVGD 161  
DB 481 QAOETILAAADLATVDTQHLGTERRELRDELKIRSCDQFRKRDQN-----LEMIG-- 532  
QY 162 EESSILYAPVQLPLVGEVYTFEFGQSERGELISQKSMLTYYAQRRAIOFKMENAKR 221  
DB 533 -----GSLMLYDQQAALKEQLSQ--EMVNSSVVSQRLA-----NSSK 569  
QY 222 CVTQDLRLSALVSTKCHSLGSSSTNFGFAKSLTRVENALVHLSGILAPKAEKATVEOE 281  
DB 570 SIEMLKEMCA-----QSLDDQ-----TNLNNKLI-----GEVMMKISDQ 602  
QY 282 VAESSYSEBELPSHMDTKHIERIPMASEQAOTVSQLHAGNLSELGNLNNMNRDLAFHL- 340



CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 1131 AA:

Query Match 5.1%; Score 120; DB 22; Length 1131;  
Best Local Similarity 22.2%; Pred. No. 0.86;  
Matches 77; Conservative 60; Mismatches 122; Indels 88; Gaps 18;

QY 8 IEQLS--KPLSDSICGVYKLEKSAFRLPNEFNVAOTLKRISQNPASDERDALQDAC 65  
DB 368 VKELSFAPMSDRSGDISREIDVASCRIKGSYR---ALPKYQOPKSGRAICREV 423  
OY 66 LN-----KWKILSDLYQFSKTTTIDILISFVAQFLDITLTSAANSLEWLADLS 118  
DB 424 LNDTWSPFMS--EDSTFVSSKTPYE--BQLHRCDEDEFLDVLNTATITVLSVQ 480  
OY 119 EKHMHLNPVLP--VETLKSDDKGKERQADAKVAFOLVGDSE--SSILYAVLQ 173  
DB 481 KK---LSRNAPEDQEKFRLLDSLGGTSEVIQR--RAIYRIGOKAPEITIESLKKNPVTA 534  
OY 174 LPLV-----GEVTFPFOQAERK---GEISOLKSMLTYYAQRFAIOFKMENAKRC 222  
DB 535 VPVVLKRLKAKEEREAQOGFNKIMREOYEKAYLKS-----DQAVNFKQNDYR-- 585  
OY 223 VTQLDRSLAVSTKCHSLGSGSTNFGAKSLITRVENAL-----VHLSGIKLAPKAERKT 277  
DB 586 -----ALRS-----KSLNLETESVYDEHQBQHSBGASAPPSAHL 621  
OY 278 V---EQEVAESSYSEGLPSHMDTKHIERIPMASEQAQ--TVSQHLH 319  
DB 622 IFVYEDRQIILEDAAA-----LISYVYKQPAIQKEDGTIHLH 661

RESULT 15  
ABG20153  
ID ABG20153 standard; Protein; 1497 AA.  
XX  
XX ABG20153:

DT 18-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #20144.  
DE  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF

XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Dermanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR  
XX  
XX N-PSDB; AAS84340.  
XX  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

Claim 20; SEQ ID NO 50512; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 1497 AA:

Query Match 5.1%; Score 120; DB 22; Length 1497;  
Best Local Similarity 21.9%; Pred. No. 1.3;  
Matches 114; Conservative 73; Mismatches 181; Indels 152; Gaps 27;

QY 26 KLEKSAFRLPNEFN--VAQTALRLKLSQNPASDERDALQELCKWKILSDLYQFSKTT 84  
DB 493 QLRHDTSSVROELEAVAGAVLSSPGSPGAVGAEOQT-----ALLRQYSEL 540  
OY 85 RDIELISFVAQAQFLDITLTSAANSLEWLADL-----SEKIMHDLNPVLPVETL 134  
DB 541 REND--DLFWMPRIVNTSLATEGG---LVDMGSKYPRHEVPDPSKPSHLK---DTV 589  
OY 135 KSDS---DKGK--ERDQADAKV---KAPFOLVGDEESSILYAPVLO-----LP--LV 177  
DB 590 RGEPNLSLEDKGRISGRNROGRVTLGPTOVFLAKKDSGQIYIVSNMKAQOEFLPKKIS 649  
OY 178 GEVTFPFOQAERKEISOLKSMLTYYAQRFAIOFKMENAK-----RCVTQDRL 229  
DB 650 GDMRGIOYKWKSESSEGRBVK-----KDSFHKLMKMDKOSTIEMELRKVTSIAEE 702  
OY 230 SALVSTKCHSLGSGSTNFGAKSLITRVEN-----ALVHLSGIKLAPKAERKVBQVA 283  
DB 703 AALDSSE-----LLTNMEDDDDTDTLTSLN-----EIAELNOQLN 739  
OY 284 ESSVSEGLPSHMDTK-----HIERIPMASEQAQTVSOHLHAGNLSLGNLNNNRD 335  
DB 740 DDSVGLAELPSSMDTEPPGARAFISKVPPGSNATPOV--EHLGTG--LKELPDVGSSDS 797  
OY 336 LAFHLREVSDYFROSEPHSPISFLLEKAIWGYLSPELLREMSBONDALSTIFNAA 395  
DB 798 ISPLLHLLEDDDDFSENEKQ-----LAEPA-----SEPDLVKNVIDEIKDFL--LSNKK 844

```
OY 396 GLNHLDOVLEPEV-----SPTVGIESPQTPOAKP-----SYSDPRS 432
      : | : | | | |
Db 845 SYGMRKEYFWPCRRARIGFSVRDAMPSTQOGLTPAEFPNRSRRRRSCRCQCSRWPRS 904
      : | | | | |
OY 433 -----VEHVQSQTSVPDTQSKODKPOSSATSALSW 463
      : | | | | |
Db 905 PTFIFTSCESRSRSH-NWSSPOELQ-KPNRAMASSNRALRSW 942
      : | | | | |
```

Search completed: March 2, 2003, 03:24:50  
Job time : 77 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 02:21:48 ; Search time 24 Seconds  
(without alignments)  
567.617 Million cell updates/sec

Title: US-09-915-706A-2

Perfect score: 2343  
Sequence: 1 MPLSKHQIEQLSKPLSDSI.....DTQSKDQKPGSSATSLSW 463

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 152.5 | 6.5         | 1503   | 4     | US-08-976-255-14  |
| 2          | 126   | 5.4         | 610    | 4     | US-09-336-447A-11 |
| 3          | 120.5 | 5.1         | 999    | 2     | US-08-770-301A-11 |
| 4          | 120.5 | 5.1         | 999    | 3     | US-09-175-581-1   |
| 5          | 119.5 | 5.1         | 1066   | 4     | US-09-541-782-8   |
| 6          | 119.5 | 5.1         | 1066   | 4     | US-09-723-820-8   |
| 7          | 114.5 | 4.9         | 955    | 1     | US-08-006-676B-1  |
| 8          | 114.5 | 4.9         | 955    | 1     | US-08-282-845-2   |
| 9          | 114.5 | 4.9         | 955    | 2     | US-08-428-414A-3  |
| 10         | 114.5 | 4.9         | 955    | 2     | PCR-US94-00324-1  |
| 11         | 114   | 4.9         | 1939   | 4     | US-09-310-187A-1  |
| 12         | 110.5 | 4.7         | 675    | 1     | US-08-317-522A-9  |
| 13         | 110.5 | 4.7         | 675    | 1     | US-08-439-818A-9  |
| 14         | 110.5 | 4.7         | 675    | 2     | US-08-751-965-9   |
| 15         | 110.5 | 4.7         | 675    | 2     | US-08-738-975-9   |
| 16         | 110.5 | 4.7         | 675    | 2     | US-08-728-626-9   |
| 17         | 110.5 | 4.7         | 675    | 3     | US-08-808-599A-9  |
| 18         | 106.5 | 4.5         | 715    | 2     | US-08-849-212-6   |
| 19         | 106.5 | 4.5         | 1234   | 4     | US-09-592-054-8   |
| 20         | 106.5 | 4.5         | 2409   | 6     | 5180808-2         |
| 21         | 106   | 4.5         | 1130   | 2     | US-08-519-547A-6  |
| 22         | 106   | 4.5         | 1481   | 2     | US-08-616-844-40  |
| 23         | 106   | 4.5         | 1481   | 2     | US-08-599-654-40  |
| 24         | 106   | 4.5         | 1481   | 3     | US-08-944-868A-40 |
| 25         | 106   | 4.5         | 1481   | 3     | US-08-944-423A-40 |
| 26         | 106   | 4.5         | 1481   | 3     | US-08-944-496-40  |
| 27         | 105.5 | 4.5         | 1312   | 2     | US-08-687-080-51  |

|    |       |     |      |   |                     |                   |
|----|-------|-----|------|---|---------------------|-------------------|
| 28 | 105.5 | 4.5 | 1886 | 4 | US-08-938-105-3     | Sequence 3, Appl1 |
| 29 | 105.5 | 4.5 | 3696 | 4 | US-09-134-001C-5080 | Sequence 5080, Ap |
| 30 | 105   | 4.5 | 2285 | 4 | US-09-308-375-2     | Sequence 2, Appl1 |
| 31 | 104.5 | 4.5 | 534  | 4 | US-09-103-664A-2    | Sequence 9, Appl1 |
| 32 | 104.5 | 4.5 | 756  | 4 | US-09-085-199B-9    | Sequence 2, Appl1 |
| 33 | 104.5 | 4.5 | 1312 | 2 | US-08-592-126-148   | Sequence 148, App |
| 34 | 104.5 | 4.5 | 2137 | 4 | US-09-134-001C-4463 | Sequence 4463, Ap |
| 35 | 104   | 4.4 | 2101 | 1 | US-08-466-390-4     | Sequence 4, Appl1 |
| 36 | 104   | 4.4 | 2101 | 1 | US-08-470-950-4     | Sequence 4, Appl1 |
| 37 | 104   | 4.4 | 2101 | 1 | US-08-467-781-4     | Sequence 4, Appl1 |
| 38 | 104   | 4.4 | 2101 | 1 | US-08-195-487-4     | Sequence 4, Appl1 |
| 39 | 104   | 4.4 | 2101 | 2 | US-08-483-924-4     | Sequence 4, Appl1 |
| 40 | 104   | 4.4 | 2101 | 4 | US-09-452-294-1     | Sequence 1, Appl1 |
| 41 | 104   | 4.4 | 2101 | 5 | PCR-US93-06160-4    | Sequence 2, Appl1 |
| 42 | 104   | 4.4 | 2293 | 4 | US-09-368-590-2     | Sequence 4, Appl1 |
| 43 | 104   | 4.4 | 2482 | 1 | US-08-328-254-6     | Sequence 6, Appl1 |
| 44 | 104   | 4.4 | 3248 | 1 | US-08-353-700-1     | Sequence 1, Appl1 |
| 45 | 104   | 4.4 | 3248 | 5 | PCR-US95-16216-1    | Sequence 1, Appl1 |

## ALIGNMENTS

RESULT 1  
US-08-976-255-14  
; Sequence 14, Application US/08976255  
; Patent No. 6136581  
; GENERAL INFORMATION:  
; APPLICANT: Jono, Keith E.  
; APPLICANT: Plozman, Gregory  
; TITLE OF INVENTION: KINASE GENES AND USES  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FASTSEQ for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,255  
; FILING DATE: No. 6136581ember 21, 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/031,675  
; FILING DATE: No. 6136581ember 22, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEEX: 67-3510  
; INFORMATION FOR SEQ. ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1503 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
; US-08-976-255-14  
Query Match 6.5%; Score 152.5; DB 4; Length 1503;  
Best Local Similarity 21.4%; Pred. No. 4e-05;  
Matches 122; Conservative 75; Mismatches 189; Indels 183; Gaps 25;

[illegible]

RESULT 2  
 US-09-336-447A-11  
 ; Sequence 11: Application US/09336447A  
 ; Patent No. 6310190  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HANSEN, ERIC J.  
 ; APPLICANT: AEBI, CHRISTOPH  
 ; APPLICANT: COPE, LESLIE D.  
 ; APPLICANT: MACIVER, ISOBEL  
 ; APPLICANT: FISKE, MICHAEL J.  
 ; APPLICANT: FREDENBURG, ROSS A.  
 ; TITLE OF INVENTION: USP11 AND USP21 ANTIGENS OF MORAXELLA CATARRHALIS  
 ; FILE REFERENCE: AMCY.024  
 ; CURRENT APPLICATION NUMBER: US/09/336,447A  
 ; CURRENT FILING DATE: 1999-06-21  
 ; NUMBER OF SEQ ID NOS: 98  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 11  
 ; LENGTH: 610  
 ; TYPE: PRT  
 ; ORGANISM: Moraxella catarrhalis  
 US-09-336-447A-11

|                       |        |                   |        |                 |
|-----------------------|--------|-------------------|--------|-----------------|
| Query Match           | 5.48;  | Score 126;        | DB 4;  | Length 610;     |
| Best Local Similarity | 20.88; | Pred. No. 0.0029; |        |                 |
| Matches               | 82;    | Conservative      | 71;    | Mismatches 166; |
|                       |        |                   | Indels | 76;             |
|                       |        |                   | Gaps   | 15              |

```

OY      43 QTALEKRLSQNPSPADERDALQ--EACLNKWKKL-----SBSLYEQFSKTRIDLEL--- 89
      |::||| |::|:|::|| |::|| |
Db      37 QDSISKLVQ---DDIDITLKQDQCKMKNYLLINQLANTLITDELNNNVIKTNTSIEALGD 92

```

```

0Y  90 -ISMVAAQFLLDTTLESAANSLFEMDLADSKHMDHLPVLRYETLKSDDCKGEROAD 148
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  93 EIGWLENDIADLESGVEELTRKNQMTLIEKDEH-----DRLAQNQAD 135

0Y  149 AK-----VKAFFOLVGD--SEESSIIYAPVLOJPLVGEYTFEDFQSAERKGEISOLKSM 200
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  136 IOTLENNVVEELFNLSGILLDQEADI----AKNNA5IEELYDPDNEVAERIGELHAYTEE 191

0Y  201 LTTTVAQERFAIOERKEMNAKRCVTOJDLRLSALYSTKCHSL---GSO5TNGFAKSLILTR 236
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  192 VNKT-----LENLITSVSKNTONIDRKRAKDINNINHIYELAAQODQDSSDITKLKN 244

0Y  237 VEMNLVHISGRIKAPKAFAKTVBEVNAFSSVSEBELSHMDTKIEKTIIPASDAQVVSQ 316
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  245 VEEGLELISGHLIDQKADL-TRDKIKALESNVEEBIL--DL5GRLLDQADULYDKIKALES 301

0Y  317 HLHGNNSELGINNMKNMDLA-----PHLREVSVDYFROSEPHSPISLELKAIRWG 368
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  302 NVEBELDL5GRLLDQKADIAQNOTDIODLAAYVELDQYIAQOKOTEAIDA--LKKASSEN 335F

0Y  369 YLSPEL-----LREMM5QNGDAISTIFNAGLN 398
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  360 TONIEDLAAYVELDQYIAQOKOTEAIDLKNAKSS5N 394

```

```

US-08-770-301A-1
US-08-770-301A-1
Sequence 1, Application US/08770301A
Patent No. 5948637
GENERAL INFORMATION:
APPLICANT: IKEDA, JUN
APPLICANT: KANEDA, SUDOIRO
APPLICANT: YANAGI, HIDEKI
APPLICANT: MATSUMOTO, MASAYASU
APPLICANT: YURA, TAKASHI
TITLE OF INVENTION: NOVEL STRESS PROTEINS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,301A
FILING DATE: 20-DEC-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1422-287
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)-205-8000
TELEFAX: (703)-205-8050
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 999 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-770-301A-1

```

|             |                                      |
|-------------|--------------------------------------|
| Query Match | 5.18; Score 120.5; DB 2; Length 999; |
|-------------|--------------------------------------|

Best Local Similarity 22.2%; Pred. No. 0.021;  
Matches 89; Conservative 63; Mismatches 122; Indels 127; Gaps 19;

QY 81 SKTRDIELISWFAQFLDITLESAA NSLEWLADLSEKHWDLNPLPVETLKSDDK 140

```
Db 690 ARRRRVEEIGVLLVLDLPDLPEDKLAQSVOKLODL-----TLR--DLE 732
Qy 141 GKREEDADKAKVAFQVLDVDSSESSILVAPVLOPLVGEVTFDFPOSARKEGISOLKSM 200
Db 733 KQEREAANSLEAFI-----FETQDKLYOPEYO-----EVS--TEEOREEISGLSASATW 781
Qy 201 L-----TTVAOERFA-----IOPKMERAKRCVQLDRLSALVSTKCHLSQSST 245
Db 782 LDEEGVATTMLKEKLAELRKLCQGLFFRVERERKKWP---EKLSDLNLDLNS-----832
Qy 246 NFGFAKSLTLRVENALVHLSGIKLAPKABA--KTYOEVAESSVSGELPSHMDTKHIER 303
Db 833 -----SMFLKGARLIPEMDQIFTEVEMTLLEKVINETWAMKNTATLAEQAK 877
Qy 304 IPMASQAOQTVSOHLHAGNLSSELGNLNNNRDLAFHLREVSDFRKQSEPHSIFLLEK 363
Db 878 LP-ATEKPVLLSKDIEAKMA-----LDREV-----QYLK 908
Qy 364 AIRWGLSLPELLREMSQNGDALSTIFNAGLNHLDOVLLPEVSTPTVGIESPOTPOA 423
Db 909 A-----KFTKP---RPRPKKNGTRAEPLNLSASDQGEKVIIP-----AGQTEDA 951
Qy 424 KPSVSDPSRVEEHVSQTSVPDQO-----SKODKPOSS 456
Db 952 EP-ISEPEKVE---TGSEPGDTPEPLGPGAEPEQKEQST 988
```

## RESULT 4

US-09-175-581-1  
Sequence 1, Application us/09175581  
Patent No. 6034232

## GENERAL INFORMATION:

APPLICANT: IKEDA, JUN  
APPLICANT: KANEDA, SUMIKO  
APPLICANT: YANAGI, HIDEKI  
APPLICANT: MATSUMOTO, MASAYASU  
APPLICANT: YURA, TAKASHI  
TITLE OF INVENTION: NOVEL STRESS PROTEINS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/175,581  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/770,301  
FILING DATE: 20-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR, GERALD M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1422-287  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)-205-8000  
TELEFAX: (703)-205-8050  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 999 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-175-581-1

Query Match 5.1%; Score 120.5; DB 3; Length 999;  
Best Local Similarity 22.2%; Pred. No. 0.021;  
Matches 89; Conservative 63; Mismatches 122; Indels 127; Gaps 19;

```
Qy 81 SKTRDIELISWFAAQFLDPTLESANSLLEMLADLSEKHMDHLPVLPVETLKSDDK 140
Db 690 ARRRRVEEIGVLLVLDLPDLPEDKLAQSVOKLODL-----TLR--DLE 732
Qy 141 GKREEDADKAKVAFQVLDVDSSESSILVAPVLOPLVGEVTFDFPOSARKEGISOLKSM 200
Db 733 KQEREAANSLEAFI-----FETQDKLYOPEYO-----EVS--TEEOREEISGLSASATW 781
Qy 201 L-----TTVAOERFA-----IOPKMERAKRCVQLDRLSALVSTKCHLSQSST 245
Db 782 LDEEGVATTMLKEKLAELRKLCQGLFFRVERERKKWP---EKLSDLNLDLNS-----832
Qy 246 NFGFAKSLTLRVENALVHLSGIKLAPKABA--KTYOEVAESSVSGELPSHMDTKHIER 303
Db 833 -----SMFLKGARLIPEMDQIFTEVEMTLLEKVINETWAMKNTATLAEQAK 877
Qy 304 IPMASQAOQTVSOHLHAGNLSSELGNLNNNRDLAFHLREVSDFRKQSEPHSIFLLEK 363
Db 878 LP-ATEKPVLLSKDIEAKMA-----LDREV-----QYLK 908
Qy 364 AIRWGLSLPELLREMSQNGDALSTIFNAGLNHLDOVLLPEVSTPTVGIESPOTPOA 423
Db 909 A-----KFTKP---RPRPKKNGTRAEPLNLSASDQGEKVIIP-----AGQTEDA 951
Qy 424 KPSVSDPSRVEEHVSQTSVPDQO-----SKODKPOSS 456
Db 952 EP-ISEPEKVE---TGSEPGDTPEPLGPGAEPEQKEQST 988
```

## RESULT 5

US-09-541-782-8  
Sequence 8, Application us/09541782  
Patent No. 6284480

## GENERAL INFORMATION:

APPLICANT: Nislow, Corey  
APPLICANT: Sakowicz, Roman  
APPLICANT: Beraud, Christophe  
TITLE OF INVENTION: Antifungal Assay  
FILE REFERENCE: 1015  
CURRENT APPLICATION NUMBER: US/09/541,782  
CURRENT FILING DATE: 2000-04-03  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 1066  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-541-782-8

Query Match 5.1%; Score 119.5; DB 4; Length 1066;  
Best Local Similarity 19.1%; Pred. No. 0.03;  
Matches 96; Conservative 78; Mismatches 161; Indels 167; Gaps 21;

```
Qy 5 KQIEQLSKPLSDSIC-----GVYLKLESAERPLNENNAQTALRKLSQNPASDER- 58
Db 373 KEYTEEDK-LKRDLMARDKNGIYLA-----EETVGEITVLKLESQREINEMK 420
Qy 59 ----DAQOACLNKWKILSD---SLYQO-----FSKTRDIE 88
Db 421 LILKALKDELQNEKIFSEVSMVLKTEQELKKTENLNTKGTLLITFKVILKTRRRK 480
Qy 89 LISMFVAQFLDPTLESANSLLEMLADLSEKHMDHLPVLP-----VETLKSDDKRE 143
Db 481 EKKEELVASHHKTRQVLTQAGQELAAADLATDTHQHLGHTIERRELDEKIRSCQGFND 540
Qy 144 REQADAKVKAFFQVLDVDSSESSILVAPVLOPLVGEVTFDFPOSARKEGISOLKSM 203
Db 541 RMQDN-----LEMIG-----GSLNLYQDQQAALKELQLSQ--EMVNS 574
```



Db 762 EATAAAMSAEODR-ENTRATLEQOULRSEERAAELASQLESTTAAKMSAEODRESTRAT 820

QY 249 FAKSLILRVNA-----LVHSGIKLAPKAEK-----TVEQEVAESVSEGEPLSHMDT 298

Db 821 LEQOULRSEERAAELASQLESTTAAKMSAEODRESTRATLEQOULRSEERAAELASQLES 880

QY 299 KHERIPMASEQAO-----TVSQHL 318

Db 881 TTAAK--MSAEODRESTRATLEQOUL 903

RESULT 8

US-08-282-845-2

; Sequence 2, Application US/08282845

; Patent No. 5719263

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; TITLE OF INVENTION: A 230kd Antigen Present in Leishmania

; TITLE OF INVENTION: Species

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Immunex Corporation

; STREET: 51 University Street

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: Apple Macintosh

; SOFTWARE: Microsoft Word for Macintosh 5.1a

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/282,845

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/006,676

; FILING DATE: JANUARY 15, 1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Perkins, Patricia Anne

; REGISTRATION NUMBER: 34,693

; REFERENCE/DOCKET NUMBER: 5004-A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206)587-0430

; TELEFAX: (206)233-0644

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 955 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-282-845-2

Query Match 4.9%; Score 114.5; DB 1; Length 955;

Best Local Similarity 20.8%; Pred. No. 0.074;

Matches 80; Conservative 56; Mismatches 150; Indels 99; Gaps 13;

QY 17 DDISICGVYKLKESAFPLRNE-----FNVAQTALRKLSONPSADERDALQEACLKWK 70

Db 535 DAELASERKLESTVAQLERQRRERVALDALQTHQRKLDQALSSERTAAER----- 587

QY 71 ILSDSLYEQES-----KTRDIELISM----- 92

Db 588 ---DQLLOQLTELOSERTQLSQVVTDRERLTRDQRIQYEGETELARDVALCAQEMEA 644

QY 93 -FVAQFLDPTTESAANSLEMLADLSEKHMDHLPVLPVETLKSDDDKGERQADAKV 151

Db 645 RYHAAYVHLQTLLELAT---EMEDALRERALAERDEAAALDAASTSONARESACERL 701

QY 152 KAFQVLGDSESSILYAPVQLPLVGEVT-----FFDQSAERKGEISQLKSM 201

Db 702 TSLEQULRESEERAAELASQLEATAAKSSAEODRENTATLEQOULRESEERAAELASQ 761

QY 202 TTVAOERFALQKMEKAKCVNQ-----LDRSALVSRKCHSIGS-----QSTNFG 248

Db 762 EATAAAMSAEODR-ENTRATLEQOULRSEERAAELASQLESTTAAKMSAEODRESTRAT 820

QY 249 FAKSLILRVNA-----LVHSGIKLAPKAEK-----TVEQEVAESVSEGEPLSHMDT 298

Db 821 LEQOULRSEERAAELASQLESTTAAKMSAEODRESTRATLEQOULRSEERAAELASQLES 880

QY 299 KHERIPMASEQAO-----TVSQHL 318

Db 881 TTAAK--MSAEODRESTRATLEQOUL 903

RESULT 9

US-08-428-414A-3

; Sequence 3, Application US/08428414A

; Patent No. 5912166

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

; TITLE OF INVENTION: LEISHMANIASIS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PC-DOS/MS-DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/428,414A

; FILING DATE: 21-Apr-1995

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: Kadlecsek, Ann T.

; REGISTRATION NUMBER: 39,244

; REFERENCE/DOCKET NUMBER: 210121.407

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; TELEX: 3723836 SEEDANDBERRY

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 955 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; US-08-428-414A-3

Query Match 4.9%; Score 114.5; DB 2; Length 955;

Best Local Similarity 20.8%; Pred. No. 0.074;

Matches 80; Conservative 56; Mismatches 150; Indels 99; Gaps 13;

QY 17 DDISICGVYKLKESAFPLRNE-----FNVAQTALRKLSONPSADERDALQEACLKWK 70

Db 535 DAELASERKLESTVAQLERQRRERVALDALQTHQRKLDQALSSERTAAER----- 587

QY 71 ILSDSLYEQES-----KTRDIELISM----- 92

Db 588 ---DQLLOQLTELOSERTQLSQVVTDRERLTRDQRIQYEGETELARDVALCAQEMEA 644

QY 93 -FVAQFLDPTTESAANSLEMLADLSEKHMDHLPVLPVETLKSDDDKGERQADAKV 151

Db 645 RYHAAYVHLQTLLELAT---EMEDALRERALAERDEAAALDAASTSONARESACERL 701

QY 152 KAFQVLGDSESSILYAPVQLPLVGEVT-----FFDQSAERKGEISQLKSM 201

Db 702 TSLBOQLRESEERAAELASOLEATAAKSSAEODRENTATLEOQLRESEERAAELASOL 761  
QY 202 TTYVAOERFATOFKEMNAKRCVTO-----LDRLSALVSTCHSGS-----OSTNFG 248  
Db 762 EATAAKKMSAEODR-ENTRATLEOQLRSEERAAELASOLSTTRAAKMSAEODRESTRAT 820  
QY 249 FAKSLTTRVENA-----LVHLSGKILAPKAEAK-----TVEOEVAESSVSEGLPSHMDT 298  
Db 821 LEQOLROSEERAAELASOLESTTRAAKMSAEODRESTRATLEOQLRESEERAAELASOL 880  
QY 299 KHIERIPMASEOAO-----TVSQHL 318  
Db 881 TTAAK--MSAEODRESTRATLEOQL 903

## RESULT 10

PCT-US94-00324-1  
; Sequence 1, Application PC/TUS9400324  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven  
; TITLE OF INVENTION: Diagnosis of Leishmaniasis  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple System 7.1  
; SOFTWARE: Microsoft Word, version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/00324  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/006,676  
; FILING DATE: 15-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perkins, Patricia Anne  
; REGISTRATION NUMBER: 34,693  
; REFERENCE/DOCKET NUMBER: 5004-WO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 955 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US94-00324-1

Query Match 4.9%; Score 114.5; DB 5; Length 955;  
Best Local Similarity 20.8%; Pred. No. 0.074;  
Matches 80; Conservative 56; Mismatches 150; Indels 99; Gaps 13;

QY 17 DSDICGVYKLEKSAFRLRNE-----FNVAQTALRKLSQNPASDERDALQOACLNWK 70  
Db 535 DAEIASERKLESTVAQERQREREVALDALQTHQRKLQELAESERTAAER----- 587  
QY 71 ILDSLVYQFS-----KTRRDIELISW----- 92  
Db 588 ---DQLLQQLTLOSETQSVVTDRELRTRDQRIQYEGETELARDVALCAQEMEA 644  
QY 93 -FVAAGFLDITLESAAASLEMIADLSEKHMHLNPVLPVETILKSDDKKEREADAKV 151  
Db 645 RYHAAVFHLQTLLELAT---EMEDALRERALARDEAAAAAELDAASTSQNARREACRL 701

QY 152 KAFVOLGDSSESSITLAPVQLPLVGEVT-----FFDQSAERKGEISQLKSM 201  
Db 702 TSLBOQLRESEERAAELASOLEATAAKSSAEODRENTATLEOQLRESEERAAELASOL 761  
QY 202 TTYVAOERFATOFKEMNAKRCVTO-----LDRLSALVSTCHSGS-----OSTNFG 248  
Db 762 EATAAKKMSAEODR-ENTRATLEOQLRSEERAAELASOLSTTRAAKMSAEODRESTRAT 820  
QY 249 FAKSLTTRVENA-----LVHLSGKILAPKAEAK-----TVEOEVAESSVSEGLPSHMDT 298  
Db 821 LEQOLROSEERAAELASOLESTTRAAKMSAEODRESTRATLEOQLRESEERAAELASOL 880  
QY 299 KHIERIPMASEOAO-----TVSQHL 318  
Db 881 TTAAK--MSAEODRESTRATLEOQL 903

## RESULT 11

US-09-310-187A-1  
; Sequence 1, Application US/09310187A  
; Patent No. 6358751  
; GENERAL INFORMATION:  
; APPLICANT: Benichou, Gilles  
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac  
; TITLE OF INVENTION: Graft Rejection  
; FILE REFERENCE: UCSF-090  
; CURRENT APPLICATION NUMBER: US/09/310,187A  
; CURRENT FILING DATE: 1999-05-12  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1939  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-310-187A-1

Query Match 4.9%; Score 114; DB 4; Length 1939;  
Best Local Similarity 21.9%; Pred. No. 0.26;  
Matches 93; Conservative 67; Mismatches 137; Indels 128; Gaps 20;

QY 7 QIEQLSKPL-----SDSICGVYK--LEKSAFRLRNE-----FNVAQTALR 47  
Db 1387 ELEEKKKKLAQRLQDAEVAEAVNAKCSLEKTKHR-LQNEIEDLMDVRSNMAAALD 1445  
QY 48 KLSQNPASDERDALQOACLNWK-----ILSDIVEGFSKTRDIELISWFAAOFL 99  
Db 1446 KKQRN-----FDKTLAEKKQKYEESQSELSQKEARSLSTELFKLNAYEESLEH 1496  
QY 100 LDTTLESAAASLEMIADLSE-----KHMHLNPV--LPVETLK-----SDDD 139  
Db 1497 LETFKREKNKLQELISDLTEQLGEGKNVHLEKVRKQLEVEKLELQSLBEARASLEHE 1556  
QY 140 KGR-----EREQADAKVKAFFQVLGDSSEST-----LAPVQL 174  
Db 1557 EGGLELRAQLEFNQIKAEIERKLAEKDEMEQAKRNHRQVDSLOTSLDAETRSRNEVLR 1616  
QY 175 --PLVGEVTFPDFO-----SAPRKEISQLKSMLETTVAQ-----ERFAT- 212  
Db 1617 KKKMEGDLNEMEIQLSIANRRAAEORQVKSLSQSLAKDTQIQLDDAVARANDLKENLAIY 1676  
QY 213 -----QFKMNAKRCVTOQLDRLSAL-----VSTRCHSLGSSQSTNFGAKSLTRVE 258  
Db 1677 ERRNNLLQAELELBRAYVEQJERSKILAEQELIFESERVQVLHQSNSLINQK--KME 1733  
QY 259 NALVHL-----SGIKLAPKAEKTYVEOEVAESSVSEGLPSHMDT-KHIERIPMASEOAO 312  
Db 1734 SDLTQLOSEVEBAVOECHNAEEK-AKKAITDAAMAEELEKKEQDTSAHLERKKKMEQTI 1792  
QY 313 TVSQHL 317  
Db 1793 KDLQH 1797



Db 346 ENT 348

```
RESULT 14
US-08-751-965-9
; Sequence 9, Application US/08751965
; Patent No. 5858360
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; City: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,965
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 12-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 675 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-751-965-9
```

Query Match 4.7%; Score 110.5; DB 2; Length 675;  
Best Local Similarity 21.2%; Pred. No. 0.1;  
Matches 77; Conservative 53; Mismatches 130; Indels 103; Gaps 18;

```
QY 127 PVLPEVETLKSDDDKGRREGADAKVKAFFOLVDSSESSIIYAVLQPLVGEVTFPFDQ 186
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 56 PLPLKLRLKRPDPSPVPEETDMD-----PLOSPTSQKDTPOIS 96
QY 187 SAERK-----GEISQKSMITTTVAOERFAIQ---FKME-----NAKR---CVTOLDR 228
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 97 SGVQKEQPLPGETITRLGVMAVAQVERKLEAQMRLITLEGRTGTEKKIADCKEYAVE 156
QY 229 LSAIVSTRKCHSLGQSNTFGAKSLITRVENALVHLSG--IKLAPKAERKVEQVEVAESS 286
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 157 FANHLESKRWVVLGTLQDEYGIQQRRLNMENTLKNRNFWILRLPPG-----202
QY 287 VSEGEIASHMDTKHIERIPMASEQAQTYSOHLHAGNLSELGNLNNMRDLAFHLIREVSD 346
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 203 -SNEEVP-----KVPYTFDD---VAVHF---SEQEGNLSEWQKELYKNYMR--GN 244
QY 347 YFROSEPHSPISFLLEKAIKRGY--LSLPELLREM-----MSQNDALSTIFNAGLN 398
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 245 Y-----ESLVSMDYAIKSPDLMSQMERGERPTMQDESEGETPTDPSAA 290
QY 399 HLDQVLLPEVSTPTVIGIESPOTPAKPSVSDPRKSVVEHVSTQTS---PVDTQSKOD-QKP 453
| : : : : : | : : : : : | : : : : : | : : : : : |
```

Db 291 HDGIVIKIEVQTNDEGSSELETPE--PLMG---QVEEHGFQDSBLGXPCEQPDLDQEP 345  
QY 454 QSS 456  
Db 346 ENT 348

```
RESULT 15
US-08-738-975-9
; Sequence 9, Application US/08738975
; Patent No. 5880267
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; City: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,975
; FILING DATE: herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 05-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-8949
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 675 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-738-975-9
```

Query Match 4.7%; Score 110.5; DB 2; Length 675;  
Best Local Similarity 21.2%; Pred. No. 0.1;  
Matches 77; Conservative 53; Mismatches 130; Indels 103; Gaps 18;

```
QY 127 PVLPEVETLKSDDDKGRREGADAKVKAFFOLVDSSESSIIYAVLQPLVGEVTFPFDQ 186
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 56 PLPLKLRLKRPDPSPVPEETDMD-----PLOSPTSQKDTPOIS 96
QY 187 SAERK-----GEISQKSMITTTVAOERFAIQ---FKME-----NAKR---CVTOLDR 228
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 97 SGVQKEQPLPGETITRLGVMAVAQVERKLEAQMRLITLEGRTGTEKKIADCKEYAVE 156
QY 229 LSAIVSTRKCHSLGQSNTFGAKSLITRVENALVHLSG--IKLAPKAERKVEQVEVAESS 286
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 157 FANHLESKRWVVLGTLQDEYGIQQRRLNMENTLKNRNFWILRLPPG-----202
QY 287 VSEGEIASHMDTKHIERIPMASEQAQTYSOHLHAGNLSELGNLNNMRDLAFHLIREVSD 346
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 203 -SNEEVP-----KVPYTFDD---VAVHF---SEQEGNLSEWQKELYKNYMR--GN 244
QY 347 YFROSEPHSPISFLLEKAIKRGY--LSLPELLREM-----MSQNDALSTIFNAGLN 398
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 245 Y-----ESLVSMDYAIKSPDLMSQMERGERPTMQDESEGETPTDPSAA 290
```

Mon Mar 3 10:44:16 2003

us-09-915-706a-2.raii

Page 9

```

OY      399  HLDVLLDPEVSTPIVGTIESPOTPOAKRVSVDPRSRVENVHSOTS-----PYDTSKOD-OKP 453
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      291  HDGIYIKIETIQTNDGSSLETFE--PLMG---QVEHGRQSDSLGAPCGEQLPDLMDPE 345
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
OY      454  QSS 456
          :|:|
Db      346  ENT 348
          :|:|

```

Search completed: March 2, 2003, 05:05:13  
Job time : 32 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 04:19:19 ; Search time 23 Seconds

(without alignments)  
759.357 Million cell updates/sec

Title: US-09-915-706a-2

Perfect score: 2343

Sequence: 1 MFLSKHQIEQLSKPLSDSI.....DTGSKQDQKPOSSATSLSW 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174566 seqs, 37721826 residues

Total number of hits satisfying chosen parameters: 174566

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications-AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID                     | Description       |
|------------|-------|-------------|--------|------------------------|-------------------|
| 1          | 134.5 | 5.7         | 1179   | 10 US-09-815-242-13262 | Sequence 13262, A |
| 2          | 126   | 5.4         | 506    | 9 US-10-157-223-7      | Sequence 11, Appl |
| 3          | 126   | 5.4         | 610    | 9 US-09-952-267-11     | Sequence 7, Appl  |
| 4          | 119.5 | 5.1         | 1179   | 10 US-09-815-242-13608 | Sequence 13608, A |
| 5          | 112   | 4.8         | 900    | 9 US-10-060-230-17     | Sequence 17, Appl |
| 6          | 110.5 | 4.7         | 900    | 9 US-10-060-230-18     | Sequence 18, Appl |
| 7          | 110.5 | 4.7         | 381    | 10 US-09-216-393-8     | Sequence 8, Appl  |
| 8          | 110.5 | 4.7         | 476    | 10 US-09-779-307-16    | Sequence 16, Appl |
| 9          | 110.5 | 4.7         | 477    | 10 US-09-779-307-4     | Sequence 4, Appl  |
| 10         | 110.5 | 4.7         | 1252   | 10 US-09-841-132-336   | Sequence 336, App |
| 11         | 109.5 | 4.7         | 824    | 9 US-09-884-001-2      | Sequence 15, Appl |
| 12         | 109   | 4.7         | 900    | 9 US-10-060-230-15     | Sequence 15, Appl |
| 13         | 109   | 4.7         | 900    | 9 US-10-060-230-16     | Sequence 16, Appl |
| 14         | 108.5 | 4.6         | 2125   | 10 US-09-919-172-29    | Sequence 29, Appl |
| 15         | 107.5 | 4.6         | 2025   | 10 US-09-815-242-5703  | Sequence 5703, Ap |
| 16         | 107.5 | 4.6         | 3158   | 10 US-09-815-242-12611 | Sequence 12611, A |
| 17         | 106   | 4.5         | 830    | 10 US-09-064-199-19    | Sequence 19, Appl |
| 18         | 106   | 4.5         | 1106   | 10 US-09-064-199-17    | Sequence 17, Appl |
| 19         | 106   | 4.5         | 1130   | 9 US-10-104-595-6      | Sequence 6, Appl  |

|    |       |     |      |                        |                   |
|----|-------|-----|------|------------------------|-------------------|
| 20 | 106   | 4.5 | 1130 | 10 US-09-064-199-18    | Sequence 18, Appl |
| 21 | 106   | 4.5 | 1207 | 10 US-09-064-199-15    | Sequence 16, Appl |
| 22 | 106   | 4.5 | 1481 | 10 US-09-371-900-40    | Sequence 40, Appl |
| 23 | 106   | 4.5 | 1481 | 10 US-09-924-417-60    | Sequence 60, Appl |
| 24 | 105.5 | 4.5 | 2437 | 10 US-09-815-242-5834  | Sequence 5834, Ap |
| 25 | 105.5 | 4.5 | 6281 | 10 US-09-815-242-12996 | Sequence 12996, A |
| 26 | 105   | 4.5 | 1138 | 10 US-09-767-215-5     | Sequence 5, Appl  |
| 27 | 105   | 4.5 | 2285 | 10 US-09-932-183A-2    | Sequence 2, Appl  |
| 28 | 104.5 | 4.5 | 1369 | 10 US-09-729-674-42    | Sequence 42, Appl |
| 29 | 104.5 | 4.5 | 1548 | 9 US-10-025-380-1095   | Sequence 1095, Ap |
| 30 | 104.5 | 4.5 | 1548 | 10 US-09-922-217-1095  | Sequence 1095, Ap |
| 31 | 104   | 4.4 | 751  | 10 US-09-864-761-38419 | Sequence 38419, A |
| 32 | 104   | 4.4 | 1210 | 9 US-10-025-380-692    | Sequence 692, App |
| 33 | 104   | 4.4 | 1210 | 10 US-09-922-217-692   | Sequence 692, App |
| 34 | 104   | 4.4 | 1210 | 10 US-09-833-265-692   | Sequence 692, App |
| 35 | 103.5 | 4.4 | 331  | 9 US-09-987-107-38     | Sequence 38, Appl |
| 36 | 103.5 | 4.4 | 2478 | 10 US-09-815-242-5816  | Sequence 5816, Ap |
| 37 | 103.5 | 4.4 | 2478 | 10 US-09-815-242-12967 | Sequence 12967, A |
| 38 | 103   | 4.4 | 1884 | 10 US-09-785-770A-17   | Sequence 17, Appl |
| 39 | 103   | 4.4 | 1907 | 10 US-09-785-770A-16   | Sequence 16, Appl |
| 40 | 103   | 4.4 | 2472 | 10 US-09-815-242-5064  | Sequence 5064, Ap |
| 41 | 102.5 | 4.4 | 396  | 10 US-09-800-729-207   | Sequence 207, App |
| 42 | 102.5 | 4.4 | 829  | 10 US-09-946-805-8     | Sequence 8, Appl  |
| 43 | 102.5 | 4.4 | 1019 | 10 US-09-801-574-76    | Sequence 76, Appl |
| 44 | 102.5 | 4.4 | 1499 | 10 US-09-911-826A-2    | Sequence 2, Appl  |
| 45 | 102.5 | 4.4 | 1945 | 9 US-09-927-597-2      | Sequence 2, Appl  |

## ALIGNMENTS

RESULT 1

US-09-815-242-13262

Sequence 13262, Application US/09815242

Patent No. US2002061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlson, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13262

LENGTH: 1179

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-815-242-13262

Query Match 5.7%; Score 134.5; DB 10; Length 1179;

Best local Similarity 21.6%; Pred. No. 0.024;



```

RESULT 4
US-09-815-242-13608
? Sequence 13608, Application US/09815242
? Patent No. US20020061569A1
? GENERAL INFORMATION:
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlsen, Karl L.
? APPLICANT: Zyskind, Judith W.
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John D.
? APPLICANT: Carr, Grant J.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OF INVENTION: Identification of Essential Genes in
? TITLE OF INVENTION: Prokaryotes
? FILE REFERENCE: ELITRA.011A
? CURRENT APPLICATION NUMBER: US/09/815,242
? CURRENT FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/289,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 13608
? LENGTH: 1179
? TYPE: PRT
? ORGANISM: Streptococcus pneumoniae
US-09-815-242-13608

```

```

: RESULT 5
: US-10-060-230-17
: Sequence 17, Application US/10060230
: Patent No. US20020173014A1
: GENERAL INFORMATION:
: APPLICANT: HILTONEN, Kaleervo
: TITLE OF INVENTION: A method to control cellular (3H)-hydroxyacyl-CoA
: TITLE OF INVENTION: esters, precursor molecules for polyhydroxyalkanoate
: TITLE OF INVENTION: synthesis in genetically modified organisms
: FILE REFERENCE: 0365-0528P
: CURRENT APPLICATION NUMBER: US/10/060.230
: CURRENT FILING DATE: 2002-02-01
: PRIOR APPLICATION NUMBER: 19991667
: PRIOR FILING DATE: 1999-08-03
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: patentIn Ver. 2.1
: SEQ ID NO 17
: LENGTH: 900
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
: US-10-060-230-17

```

```

QY 2 PLKHOIEQ-----LSPIKSDSDICGYYLLEKSAAPRLNENNAOTLRKLSQNSPA 55
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 195 PLASRMTEENVLPRLHKQLOPEKIVPVLVYLTHESIKVSNSTIELAAGFGQLRMERS 254
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 56 -----DERDALOEACLKMKWILSDSYEOSTKTRDIELIMFVAOFLDPTTLESAAN 109
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 255 GQIFNPDKRYTPPALINKKEITDYDRDKPRNKQHPHYQLSDY-----NDLIRKAK 305
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 110 SLEMLADLSEKHMHDLNPLYVEFTLKSDDDKGER-----EQADAKY-----KAFQLYGD 160
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db 306 KLPPEQGSVKIKSLCNKVYVTS--AGGGLGKSHAIWAFRGAKVYNDIKDFESV-- 361

Qy 161 SPESSILVAP-----VQLPLVGVTFPFDDQSAE---RKGISLTKSMLTITVAQ 207

Db 362 -EIKKLLGEGTAIPDSHDVTEAPILITOTALSKFQRYDILVNNAGILDKGFLKKR-DE 419

Qy 208 ERFPAI-----QFKMENAKRCVTOLDRLSALVST-----KCHLSQSQSTNFGF 249

Db 420 EEFPAVLKHLHSTFSLSKAVMPILFKQKSGFLIINTSTSGIYGNFGQAVAAAKAAILGF 479

Qy 250 AKSLITRYENALVHLGKIK-----LAPKAQKTYEQAEEVASSVSEGLPSSHMDTKHIE 302

Db 480 SKT-----IALEGARGIIVNVIAPHAET-----AMTKTISEKELSHNHFDAQSVS 525

Qy 303 R--IPMASEQAQTVSOHLHAGNLSLQ 327

Db 526 PLVYLASEELQKYSGRRVIGQLFEVG 552

```

RESULT 6
US-10-060-230-18
: Sequence 18, Application US/10060230
: Patent No. US20020173014A1
: GENERAL INFORMATION:
: APPLICANT: HILFENEN, Kaleervo
: TITLE OF INVENTION: A method to control cellular (3R)-hydroxyacyl-CoA
: TITLE OF INVENTION: esters, precursor molecules for polyhydroxyalkanoate
: TITLE OF INVENTION: synthesis in genetically modified organisms
: FILE REFERENCE: 0365-0528P
: CURRENT APPLICATION NUMBER: US/10/060,230
: CURRENT FILING DATE: 2002-02-01
: PRIOR APPLICATION NUMBER: 19991667
: PRIOR FILING DATE: 1999-08-03
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 18
: LENGTH: 900
: TYPE: prt
: ORGANISM: Saccharomyces cerevisiae
US-10-060-230-18

```

|                       |        |              |       |                 |
|-----------------------|--------|--------------|-------|-----------------|
| Query Match           | 4.8%;  | Score 112;   | DB 9; | Length 900;     |
| Best Local Similarity | 21.7%; | Pred. No. 1; |       |                 |
| Matches               | 84;    | Conservative | 59;   | Mismatches 154; |
|                       |        |              |       | Indels 90;      |
|                       |        |              |       | Gaps 16         |

```

OY      2  PLSKIOIQO-----LSRPLSDSDSICGVYLLKESAPRPLNENNAVOTARUKISQNSPA 55
Db      195 PLASRMTEENVPRIILIKOLOREKTYRVLVLLTHESTRKVSSTIELAAGFGQILRWERS 25
OY      56  -----DERDALQEAACLNKMKILSDSIYEQSKTTRDIELLSWFEVAOFLDITLLESAN 109
Db      255 GOIFNPDEKTYTPEAILNKMKREITDYRDKRPNNKOTHPYOLSDY-----NDLITKAK 30
OY      110 SLEMLADLSKXHHOHLNVLYRVEVELIKSDDKGKER-----EQADAKV-----KAFFOLYGD 166
Db      306 KLPNPQGSVAKIKSLCNKVVAVVS--AGGGLGKSHAIWFARYGAKVAVNDIKDFSVV-- 361
OY      161 SEESSILVAP-----VLQPLVEYVEFFPOSAAE--RKGEISOLKSMYLTITVAO 207
Db      362 -EIKKKLIGEGCAIPDSDHVVTEAPRILOTAISKFQKRVILVNNAGILROKSPFKMK-DE 419
OY      208 ERFAL-----QFKMENAKRCVQLODLALSVST-----KCHSLGSSOTNEGF 245
Db      420 EWFVAVLKVLHSTFSLSKAAVPIFKQKSCPIINTTSGIYGNQGNVAAAKAILGF 475
OY      250 AKSLITREYNALVHLSGK-----LAPKAEATVQDEVAESSVSGEELPSHMDTKHIE 307
Db      480 SKT-----IALEGAKRGIIYVAVAPHAET-----AMTKTIFSEKELSNHEDASOVS 525
OY      303 R--IPMASEOAOQVYQSHLHAGNLSEIG 327
Db      526 PLVYVLLASEELOKYSGRVYIQOLEEYVG 552

```

```

RESULT 7
US-09-216-393-8
: Sequence 8, Application US/09216393
: Patent No. US2001001447A1
: GENERAL INFORMATION:
: APPLICANT: Milhausen, Michael James
: TITLE OF INVENTION: TOKOPLASMA GONDI1 PROTEINS, NUCLEIC ACID MOLECULES, AND
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: TX-1-C2
: CURRENT APPLICATION NUMBER: US/09/216.393
: CURRENT FILING DATE: 1998-12-18
: EARLIER APPLICATION NUMBER: 08/994,825
: EARLIER FILING DATE: 1997-12-19
: NUMBER OF SEQ ID NOS: 364
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 381
: TYPE: PR1
: ORGANISM: Tokoplasma gondi1
US-09-216-393-8

```

|                          |        |                 |             |             |
|--------------------------|--------|-----------------|-------------|-------------|
| Query Match              | 4.7%;  | Score 110.5;    | DB 10;      | Length 381; |
| Best Local Similarity    | 21.4%; | Pred. No. 0.38; |             |             |
| Matches 87; Conservative | 44;    | Mismatches 110; | Indels 165; | Gaps 19;    |

[illegible]

RESULT 8  
 US-09-779-307-16  
 : Sequence 16, Application US/09779307  
 : Patient No. US20020137675A1  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Taupier Jr., Raymond  
 : APPLICANT: Majumder, Kundu  
 : APPLICANT: Vernet, Corinne  
 : APPLICANT: Prayaga, Sudhidas  
 : TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Therapy  
 : FILE REFERENCE: 15966-662 US  
 :  
 : CURRENT APPLICATION NUMBER: US/09/779,307  
 : CURRENT FILING DATE: 2001-02-07  
 :  
 : PRIOR APPLICATION NUMBER: 60/180,880  
 : PRIOR FILING DATE: 2000-02-08  
 :  
 : PRIOR APPLICATION NUMBER: 60/181,044  
 : PRIOR FILING DATE: 2000-02-08  
 :  
 : PRIOR APPLICATION NUMBER: 60/181,656

```

:
:
: PRIOR FILING DATE: 2000-02-10
: PRIOR APPLICATION NUMBER: 60/182,795
:
: PRIOR FILING DATE: 2000-02-15
:
: NUMBER OF SEQ ID NOS: 43
:
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ ID NO 16
:
: LENGTH: 476
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
US-09-779-307-16

```

|                       |        |                 |        |                 |
|-----------------------|--------|-----------------|--------|-----------------|
| Query Match           | 4.7%;  | Score 110.5;    | DB 10; | Length 476;     |
| Best Local Similarity | 21.8%; | Pred. NO. 0.52; |        |                 |
| Matches               | 78;    | Conservative    | 53;    | Mismatches 134; |
|                       |        |                 | Indels | 93;             |
|                       |        |                 | Gaps   | 15;             |

```

QY 27 LEKAFRLRNEFVAOTLAKRLSONSPADRLOAECKMMLTSP-----SLVEOSK 83
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 VNQSLSTLNEVPNQAMKTOEE-----QIKTLKRFSPYDKVASTLQONKML 110
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 84 TRDIELISWEYAAQFLDITLTLESANSLW-----LADLSERH 123
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 ETWKSLLOOQOATARSNNDMFESVNNRLMQLETIGOEKLTLEAELGNQPLVEDFPK 176
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 WDHLNPLVPEVT-----LKSDDDKGEREBOADAKVAFQOLYGDSEESSITVAPILOPLV 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 EDEINKRTKEMENEFVLKKNAD-----EASNNKVELSRLEGLDINFEL-----MQL--- 224
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 GEYFFPDQSKERGESOLKSMLTITTYAODERFALQIKREMAKRCVYQDLRLSLVSTKC 233
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 -----YEEELRELOQSISDT-----SVLMSDMS--CSLDMOSIIAEVNAQY 264
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 238 HSLGOSQSTNFGAKSLITRVENALVHLSCILAKRAEAKTYEODEVAESSVEGELPSHMD 297
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 265 EELNANCSR--AEASMSYQIKYEELQTLAGKHGDDRRRTK---EISELYIGNISQLQAHVD 311
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 298 TKHIERIPMASEQOYQSQHLLA-----GNLSL-GNLNNNNRDLAFHLLREVS 346
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 OR-----SLLEAITDAELHGLGVYKANAKITSELEALQIRAKODAMQ--LEEYD 369
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 9
US-09-779-307-4
; Sequence 4, Application US/09779307
; Patent No. US20020137675A1
; GENERAL INFORMATION:
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Majumder, Kumud
; APPLICANT: Vernel, Corine
; APPLICANT: Prayaga, Subhidas
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Therapy
; FILE REFERENCE: 15966-662 US
; CURRENT APPLICATION NUMBER: US/09/779,307
; CURRENT FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/180,880
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 60/181,044
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 60/181,656
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 60/182,795
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-779-307-4

```

Query Match 4.7%; Score 110.5; DB 10; Length 477;  
Best Local Similarity 21.8%; Pred. No. 0.52;  
Matches 78; Conservative 53; Mismatches 134; Indels 93; Gaps 15;

```

QY 27 LKSAFRLRLRENFVNAOTALRLKLSQNSADEROMLOZACLNKKILSD---SLYEORSKT 83
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 VNOSILSPLELVDPNIOAMRTQEE-----QIKTLNKFAFPYDKVRSLEOQMKML 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 84 TRDIELSMFEVAOFLIDPTLLESAANLEW-----LADLSEKH 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 ETKSKLSLOOQOTAKSNMNMNFESVYNNLRKQOLETGOEKLLEAEVLSNMOPLYDEFRNKY 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 WDHNLNPLVLPYET-----LKSDDDKQKEREQADAKAFQOLVGBSEESSILYAVLQPLV 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 EDELINKRTKEMENEVLILKKQAD-----EASNNKVELGSRLIEGLTDELNFL----MQL--- 225
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 GEVTFEFDQSAERKEISOLKSMYLTTPVAOERAIQKEMENAKRCYTQIDRLSALVSTKC 237
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 -----YEEETIRELOQSIDPT-----SVLSDMDS--CSLDMDSIIAEVMAQY 265
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 238 HSLGSQSTNNGFAASLLTRVBNALVHLSGIKRLAPKAARKVVEQVEVAESSVSEBELPSHMD 297
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 EELNACSR-AEASSMOIKYEELQTLAGKHGDDRRTKT--EISEIYGNISQLOAHVD 320
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 298 TKHTERIPMASEGOQVTSQHLLA-----GNLSEL-GNUNNNRRGLARHLREYSD 346
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 QR-----ASLEATIDAEHLGOLVYKQDANAKLSELEALQRLQAKQDMAQD-LREYQ 370
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
US-09-841-132-396
: Sequence 396, Application US/09841132
: Patent No. US20020061848A1
: GENERAL INFORMATION:
:
: APPLICANT: Bhalla, Ajay
: APPLICANT: Skeiky, Jasir A.W.
: APPLICANT: Probst, Peter
: TITLE OF INVENTION: COMPOSITIONS AND METHODS
: FILE REFERENCE: 210121.469C8
: CURRENT APPLICATION NUMBER: US/09/841,132
: CURRENT FILING DATE: 2001-04-23
: NUMBER OF SEQ ID NOS: 599
: SOFTWARE: FastSeq for Windows Version 3.0/4.0.
: SEQ ID NO 396
: LENGTH: 1252
:
: TYPE: PR1
: ORGANISM: Chlamydia pneumoniae
: US-09-841-132-396

```

|                       |        |                |        |                 |
|-----------------------|--------|----------------|--------|-----------------|
| Query Match           | 4.7%;  | Score 110.5;   | DB 10; | Length 1252;    |
| Best Local Similarity | 20.1%; | Pred. NO. 2.1; |        |                 |
| Matches               | 96;    | Conservative   | 76;    | Mismatches 169; |
|                       |        |                | Indels | 137;            |
|                       |        |                | Gaps   | 25;             |

[illegible]



Best Local Similarity 21.7%; Pred. No. 1.7;  
Matches 84; Conservative 58; Mismatches 155; Indels 90; Gaps 16

```

0Y 2 PLKHQIQEQ-----LSKPLSDSDSICGVYKLEKSAARPLRNNVNAQTLRKLIQNSPSA 55
Db 195 PLASRMEENVLPHILKQIQEPKIVPLVYLTHESKVSNSIFELAGFGGOLRWERS 254
0Y 56 -----DERDALOEACLNKKMILKSDSYEOPFSKTRDIELISMFVAQAQFLDPTLLESAAN 109
Db 255 GQIFNPDPKTTTPRALINKKKEITDYDKRPKNKTQHPYQLSDY-----NDLTKKAK 30505
0Y 110 SLEWLADLSEKHMDHLNPVLPVETPLKSDDDGKGER-----EQADAKV-----KAFFOLVGD 160
Db 306 KLPRNEQGSYVXIKSLCKNKVVVV--TGAGGGGLGKSHALIMPARYAKKVVYVNDIKDPFSYV-- 361
0Y 161 SEESSILYAP-----VLQPLRYGEYTFPFDQSAE--RKGEISQLKSMLTTTVAQ 207
Db 362 -EETINKLIGESTAPDSHDVYTEAPLILQTAISKFQEVLDILVNNAGLIRKSPFLKMK-DE 419
0Y 208 ERPAI-----OFKEMNAKCVQDLRLSALVST-----KCHSLGSGOSTMFGF 249
Db 420 EMFPAVLKVLHLESTSLSKAAVPIPTFKKSGEILINTTSGIYGNFGGANYAAKAAALIGF 479
0Y 250 AKSLTRVENALVHLISGIK-----LAPRAEAKTVQGEVAESSSEGLPSHMDTKHIE 302
Db 480 SKT-----LALLEGAKRRIIYNVIAPIAHET-----AMTKITFSKEKLSNHFDAQSVS 525
0Y 303 R--IPMASEQAQIVSOQLHAGNLSLQ 327
Db 526 PLVYVLLASEELOKYSGRVRVIGOLEEYVG 552

```

```

RESULT 14
US-09-919-172-29
Sequence 29, Application US/09919172
Patent No. US20020119463A1
GENERAL INFORMATION:
APPLICANT: Faris, Mary
APPLICANT: Turner, Christopher M.
TITLE OF INVENTION: PROSTATE CANCER MARKERS
FILE REFERENCE: PA-0036 US
CURRENT APPLICATION NUMBER: US/09/919,172
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/222,469
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 102
SOFTWARE: PERL Program
SEQ ID NO 29
LENGTH: 2125
TYPE: PRF
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20020119463A1 3774181C01
US-09-919-172-29

```

|                       |                  |                 |            |              |
|-----------------------|------------------|-----------------|------------|--------------|
| Query Match           | 4.6%;            | Score 108.5;    | DB 10;     | Length 2125; |
| Best Local Similarity | 20.3%;           | Pred. No. 6.6;  |            |              |
| Matches 76;           | Conservative 69; | Mismatches 159; | Indels 71; | Gaps 15;     |

[illegible]

|    |     |  |     |
|----|-----|--|-----|
| Db | 596 | SDILOROKATVLENSKLTGKISELERVAYELKKO-----KSHVEELPKVREAENELRKQ  | 651 |
| Qy | 220 | KRCVTLORLSALVSTKCHLSGOSTNGEFAKSLUTRY-----ENALVHLSGILAKPA     | 273 |
| Db | 652 | QNVEDI---SLOKIRESEAKO-----YRRELETVIVREKAERLELERYDOLTTIEAA    | 702 |
| Qy | 274 | EAKTVEQ-----EVAESSVSEBELPDSHMDTKIIRIPASBEAQVSHLAGNISE-       | 325 |
| Db | 703 | KRAAAEENLNFRRNOLEENFTPTRTLEDHLKRKDL-S-LNDLEQOKNLMEELRRKRDNEE | 761 |
| Qy | 326 | --LGNLNNNRDLAF   | 338 |
| Db | 762 | ELKLKIKOMEKDLAF  | 776 |

```

RESULT 15
US-09-815-242-5703
? Sequence 5703, Application US/09815242
? Patent No. US20020061569A1
? GENERAL INFORMATION:
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlсен, Karl L.
? APPLICANT: Zyskind, Judith W.
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John D.
? APPLICANT: Carr, Grant J.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OF INVENTION: Identification of Essential Genes In
? TITLE OF INVENTION: Prokaryotes
? FILE REFERENCE: ELTRA.011A
? CURRENT APPLICATION NUMBER: US/09/815.242
? CURRENT FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 5703
? LENGTH: 2025
? TYPE: PRT
? ORGANISM: Staphylococcus aureus
? FEATURE:
? NAME/KEY: VARIANT
? LOCATION: (1)..(2025)
? OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5703

```

|                      |       |              |       |            |     |        |      |
|----------------------|-------|--------------|-------|------------|-----|--------|------|
| Query Match          | 4.6%  | Score        | 107.5 | DB         | 10  | Length | 2025 |
| Best Local Similarly | 21.8% | Pred. No.    | 7.4   |            |     |        |      |
| Matches              | 96    | Conservative | 66    | Mismatches | 157 | Indels | 121  |
|                      |       |              |       |            |     | Gaps   | 21   |

```
OY      37 NEVNAOTALRKLSONSPAER--DLLOECLJNKWKIISDSLYE-OFTSTDEILJSMF 93
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     1185 NDINGALOAKDAIONLNGROEQLEAKDKAIQNVKVLADLKLEIASATODOKLIA-- 1242
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      94 VAAFLDITLESANSLWKLADLSKHMD-----HLNVLPEVTLSKSD 138
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     1243 -----KNKAELLANSIINNIKATSNDVOVSOTAGNOALEOVHAN----ELPKAKI 1290
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY     139 DKGERBOADAKVAFFOLG-----DSSESSILYAPVQLPLVEVFDFDOSAGERGE 193
```

```

Db 1291 DANKD---VDKQVQALIDEIGRNPLTDKEQALKDRINQILOQGH---NDINNALTKEA 1344
QY 194 ISOLKSMLTIVAGERFAIOFKMENAKRCVYQLOBRSLAVSTKCHSLGSQSTNFGFAKSL 253
Db 1345 IEQAKERLAQALQDIDIKDLVKAQ-EDAK-----NKIKALANAKKDQINSNP----- 1388
QY 254 LTRVENALVHLGKIKLAPKAEAKTVEQEVAESSVSEGELPSHMDPKHIERIPMAEQOQT 313
Db 1389 -----DLTPEQAKAKLKE-----IDEAE-----KRALQVENAQOT 1418
QY 314 VSQHLHAGNLSELGNLNNMNDLAFHLIREVSDYFRQSEPHSPISFLEKAIKMGYLSLP 373
Db 1419 IDQ-LNRGLNLGLDDIRNTH-----VWEVDD-----QPAVNEISEATPEQLLVNG 1462
QY 374 ELL---REMSEONGDALSTIFNAGLNHLDOVLLPEVSTP-TVGIESPQTPQAKPSVSD 429
Db 1463 ELIYHRDDITEQ-DVL-----AHINLIDQLAEVITDPSTATISDSLAKVEVTLLD 1514
QY 430 PRSVEEHVSQTSPYDTOSQ 449
Db 1515 GSKVIYVY---PVKVEKE 1530

```

Search completed: March 2, 2003, 05:44:03  
 Job time : 55 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 03:25:28 ; Search time 266 Seconds  
(without alignments)  
1122.223 Million cell updates/sec

Title: US-09-915-706A-2

Perfect score: 2343

Sequence: 1 MPLSKHQIQLEKSPLESDSI.....DTGSKQDQKQSSATSALSW 463

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 64473110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/paa/US06.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US081.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US082.COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US083.COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US084.COMB.pep:\*  
9: /cgn2\_6/ptodata/1/paa/US085.COMB.pep:\*  
10: /cgn2\_6/ptodata/1/paa/US086.COMB.pep:\*  
11: /cgn2\_6/ptodata/1/paa/US087.COMB.pep:\*  
12: /cgn2\_6/ptodata/1/paa/US088.COMB.pep:\*  
13: /cgn2\_6/ptodata/1/paa/US089.COMB.pep:\*  
14: /cgn2\_6/ptodata/1/paa/US090.COMB.pep:\*  
15: /cgn2\_6/ptodata/1/paa/US091.COMB.pep:\*  
16: /cgn2\_6/ptodata/1/paa/US092.COMB.pep:\*  
17: /cgn2\_6/ptodata/1/paa/US093.COMB.pep:\*  
18: /cgn2\_6/ptodata/1/paa/US094.COMB.pep:\*  
19: /cgn2\_6/ptodata/1/paa/US095.COMB.pep:\*  
20: /cgn2\_6/ptodata/1/paa/US096.COMB.pep:\*  
21: /cgn2\_6/ptodata/1/paa/US097.COMB.pep:\*  
22: /cgn2\_6/ptodata/1/paa/US098.COMB.pep:\*  
23: /cgn2\_6/ptodata/1/paa/US099.COMB.pep:\*  
24: /cgn2\_6/ptodata/1/paa/US100.COMB.pep:\*  
25: /cgn2\_6/ptodata/1/paa/US101.COMB.pep:\*  
26: /cgn2\_6/ptodata/1/paa/US102.COMB.pep:\*  
27: /cgn2\_6/ptodata/1/paa/US60.COMB.pep:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID                | Description       |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1          | 2343  | 100.0       | 463    | US-09-915-706A-2     | Sequence 2, Appl1 |
| 2          | 264.5 | 11.3        | 364    | US-09-328-352-6641   | Sequence 6641, Ap |
| 3          | 197   | 8.4         | 681    | US-09-252-981A-25690 | Sequence 25690, A |
| 4          | 152.5 | 6.5         | 1503   | US-09-649-996-14     | Sequence 14, Appl |
| 5          | 144   | 6.1         | 151    | US-09-739-449-9942   | Sequence 9942, Ap |
| 6          | 144   | 6.1         | 151    | US-09-803-110-9942   | Sequence 9942, Ap |

|  | 7  | 138   | 5.9 | 2008 | 14  | US-09-091-501-8      | Sequence 8, Appl1 |
|--|----|-------|-----|------|---|----------------------|-------------------|
|  | 8  | 138   | 5.9 | 2008 | 14 <td>US-09-091-501B-8</td> <td>Sequence 8, Appl1</td>     | US-09-091-501B-8     | Sequence 8, Appl1 |
|  | 9  | 137.5 | 5.9 | 357  | 16 <td>US-09-252-991A-17293</td> <td>Sequence 17293, A</td> | US-09-252-991A-17293 | Sequence 17293, A |
|  | 10 | 136.5 | 5.8 | 5984 | 27 <td>US-60-173-464-18560</td> <td>Sequence 18560, A</td>  | US-60-173-464-18560  | Sequence 18560, A |
|  | 11 | 136.5 | 5.8 | 7118 | 27 <td>US-60-167-217-22879</td> <td>Sequence 22879, A</td>  | US-60-167-217-22879  | Sequence 22879, A |
|  | 12 | 136.5 | 5.8 | 7118 | 27 <td>US-60-173-464-20674</td> <td>Sequence 20674, A</td>  | US-60-173-464-20674  | Sequence 20674, A |
|  | 13 | 136.5 | 5.8 | 7182 | 27 <td>US-60-191-637-39835</td> <td>Sequence 39835, A</td>  | US-60-191-637-39835  | Sequence 39835, A |
|  | 14 | 136.5 | 5.8 | 7182 | 27 <td>US-60-191-681-30862</td> <td>Sequence 30862, A</td>  | US-60-191-681-30862  | Sequence 30862, A |
|  | 15 | 136.5 | 5.8 | 7201 | 20 <td>US-09-614-150-40200</td> <td>Sequence 40200, A</td>  | US-09-614-150-40200  | Sequence 40200, A |
|  | 16 | 135   | 5.8 | 372  | 22 <td>US-09-897-516-7344</td> <td>Sequence 7344, Ap</td>   | US-09-897-516-7344   | Sequence 7344, Ap |
|  | 17 | 135   | 5.8 | 372  | 22 <td>US-60-215-161-7344</td> <td>Sequence 7344, Ap</td>   | US-60-215-161-7344   | Sequence 7344, Ap |
|  | 18 | 135   | 5.8 | 2013 | 24 <td>US-10-089-928-9</td> <td>Sequence 9, Appl1</td>      | US-10-089-928-9      | Sequence 9, Appl1 |
|  | 19 | 134.5 | 5.7 | 1179 | 1 <td>PCT-US02-03987-13262</td> <td>Sequence 13262, A</td>  | PCT-US02-03987-13262 | Sequence 13262, A |
|  | 20 | 134.5 | 5.7 | 1179 | 21 <td>US-09-752-069A-219</td> <td>Sequence 219, App</td>   | US-09-752-069A-219   | Sequence 219, App |
|  | 21 | 134.5 | 5.7 | 1179 | 22 <td>US-09-815-242-13262</td> <td>Sequence 13262, A</td>  | US-09-815-242-13262  | Sequence 13262, A |
|  | 22 | 134.5 | 5.7 | 1179 | 24 <td>US-10-072-851-13262</td> <td>Sequence 13262, A</td>  | US-10-072-851-13262  | Sequence 13262, A |
|  | 23 | 134.5 | 5.7 | 1179 | 27 <td>US-60-174-089-219</td> <td>Sequence 219, App</td>    | US-60-174-089-219    | Sequence 219, App |
|  | 24 | 133   | 5.7 | 1710 | 27 <td>US-60-242-679-970</td> <td>Sequence 970, App</td>    | US-60-242-679-970    | Sequence 970, App |
|  | 25 | 129.5 | 5.5 | 630  | 21 <td>US-09-791-537-61506</td> <td>Sequence 61506, A</td>  | US-09-791-537-61506  | Sequence 61506, A |
|  | 26 | 129.5 | 5.5 | 630  | 21 <td>US-09-791-537-76922</td> <td>Sequence 76922, A</td>  | US-09-791-537-76922  | Sequence 76922, A |
|  | 27 | 129.5 | 5.5 | 860  | 24 <td>US-10-037-417-59</td> <td>Sequence 59, Appl</td>     | US-10-037-417-59     | Sequence 59, Appl |
|  | 28 | 129.5 | 5.5 | 860  | 24 <td>US-10-080-334-166</td> <td>Sequence 166, App</td>    | US-10-080-334-166    | Sequence 166, App |
|  | 29 | 129.5 | 5.5 | 1036 | 27 <td>US-60-161-932-977</td> <td>Sequence 977, App</td>    | US-60-161-932-977    | Sequence 977, App |
|  | 30 | 129.5 | 5.5 | 1036 | 27 <td>US-60-167-217-4618</td> <td>Sequence 4618, Ap</td>   | US-60-167-217-4618   | Sequence 4618, Ap |
|  | 31 | 129.5 | 5.5 | 1036 | 27 <td>US-60-171-627-366</td> <td>Sequence 366, App</td>    | US-60-171-627-366    | Sequence 366, App |
|  | 32 | 129.5 | 5.5 | 1036 | 27 <td>US-60-173-464-3755</td> <td>Sequence 3755, Ap</td>   | US-60-173-464-3755   | Sequence 3755, Ap |
|  | 33 | 128.5 | 5.5 | 539  | 16 <td>US-09-270-767-43655</td> <td>Sequence 43655, A</td>  | US-09-270-767-43655  | Sequence 43655, A |
|  | 34 | 128   | 5.5 | 2017 | 1 <td>PCT-US01-08631-36660</td> <td>Sequence 36660, A</td>  | PCT-US01-08631-36660 | Sequence 36660, A |
|  | 35 | 127   | 5.4 | 1726 | 1 <td>PCT-US02-21179-13</td> <td>Sequence 13, Appl</td>     | PCT-US02-21179-13    | Sequence 13, Appl |
|  | 36 | 127   | 5.4 | 2278 | 1 <td>PCT-US02-22866-14</td> <td>Sequence 14, Appl</td>     | PCT-US02-22866-14    | Sequence 14, Appl |
|  | 37 | 126.5 | 5.4 | 752  | 21 <td>US-09-917-234-70</td> <td>Sequence 13635, A</td>     | US-09-917-234-70     | Sequence 13635, A |
|  | 38 | 126.5 | 5.4 | 752  | 23 <td>US-09-917-234-70</td> <td>Sequence 70, Appl</td>     | US-09-917-234-70     | Sequence 70, Appl |
|  | 39 | 126.5 | 5.4 | 752  | 25 <td>US-10-171-311-48</td> <td>Sequence 48, Appl</td>     | US-10-171-311-48     | Sequence 48, Appl |
|  | 40 | 126.5 | 5.4 | 824  | 25 <td>US-10-108-260A-3569</td> <td>Sequence 3569, Ap</td>  | US-10-108-260A-3569  | Sequence 3569, Ap |
|  | 41 | 126.5 | 5.4 | 893  | 1 <td>PCT-US02-31095-51</td> <td>Sequence 51, Appl</td>     | PCT-US02-31095-51    | Sequence 51, Appl |
|  | 42 | 126   | 5.4 | 506  | 11 <td>US-08-712-708-7</td> <td>Sequence 7, Appl1</td>      | US-08-712-708-7      | Sequence 7, Appl1 |
|  | 43 | 126   | 5.4 | 506  | 25 <td>US-10-157-223-7</td> <td>Sequence 1648, Ap</td>      | US-10-157-223-7      | Sequence 1648, Ap |
|  | 44 | 126   | 5.4 | 506  | 27 <td>US-60-360-039-1648</td> <td>Sequence 11, Appl</td>   | US-60-360-039-1648   | Sequence 11, Appl |
|  | 45 | 126   | 5.4 | 610  | 23 <td>US-09-952-267-11</td> <td></td>                      | US-09-952-267-11     |                   |

#### ALIGNMENTS

RESULT 1  
US-09-915-706A-2  
; Sequence 2, Application US/09915706A  
; GENERAL INFORMATION:  
; APPLICANT: NELSON, DAVID R.  
; TITLE OF INVENTION: A LIVE, AVIRULENT STRAIN OF V. ANGIILLARUM THAT  
; TITLE OF INVENTION: PROTECTS FISH AGAINST INFECTION BY VIRULENT V.  
; TITLE OF INVENTION: ANGIILLARUM  
; FILE REFERENCE: 5112  
; CURRENT APPLICATION NUMBER: US/09/915,706A  
; CURRENT FILING DATE: 2001-07-26  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 463  
; TYPE: PRN  
; ORGANISM: Vibrio anguillarum  
US-09-915-706A-2

| Query Match           | 100.0%         | Score 2343   | DB 23                                   | Length 463 |
|-----------------------|----------------|--|---|------------|
| Best Local Similarity | 100.0%         | Pred. No. 7.6e-205                                       |   |            |
| Matches 463           | Conservative 0 | Mismatches 0   | Indels 0                                | Gaps 0     |
| QY                    | 1              | MPLSKHQIQLEKSPLESDSI                                     | CGYYLLEKSAFRLNEFNVAOTATLRKLSQNPSPADERDA | 60         |
| DB                    | 1              | MPLSKHQIQLEKSPLESDSI                                     | CGYYLLEKSAFRLNEFNVAOTATLRKLSQNPSPADERDA | 60         |
| QY                    | 61             | LOEACLNKKKTLSDSLYEQFSKTRTDIELISFVAQAFLDTTLESANSLMLADLSRK |   | 120        |

Db 61 LOEACLNKWKILSDSYEDEFKTTTRDIELISWFAAOFLDPTLESANSLWMLADISEK 120  
QY 121 HMDHLNVLPEVETLKSDDDKEREQADAKVKAFFQVGDSEESSIYAPVLOPLVGEV 180  
Db 121 HMDHLNVLPEVETLKSDDDKEREQADAKVKAFFQVGDSEESSIYAPVLOPLVGEV 180  
QY 181 TFFDFOSEAKKEGELSOLKSMLTITVAOERPAIOFKMENAKRCVTOIDRLSALVSTKCHSL 240  
Db 181 TFFDFOSEAKKEGELSOLKSMLTITVAOERPAIOFKMENAKRCVTOIDRLSALVSTKCHSL 240  
QY 241 GSOSTNGCFKKSILTRRNALVHLSGILKPKAKAEKVEQVAVSSVSEGLPSHMDTKH 300  
Db 241 GSOSTNGCFKKSILTRRNALVHLSGILKPKAKAEKVEQVAVSSVSEGLPSHMDTKH 300  
QY 301 IERIPMAEQOQVSOHLHAGNLSLGNLNNMNDLAFHLLREVSDFROSEPSPISTFL 360  
Db 301 IERIPMAEQOQVSOHLHAGNLSLGNLNNMNDLAFHLLREVSDFROSEPSPISTFL 360  
QY 361 LEKAIKRGVLSLPELLREKMSSEONGDALSTFNAGLNHLDVLLPEVSTPTVIGIESPOT 420  
Db 361 LEKAIKRGVLSLPELLREKMSSEONGDALSTFNAGLNHLDVLLPEVSTPTVIGIESPOT 420  
QY 421 POAKPSVSDRVSVEHVSQTSPTVTOQSKODOKPOSSATSALSW 463  
Db 421 POAKPSVSDRVSVEHVSQTSPTVTOQSKODOKPOSSATSALSW 463

RESULT 2  
US-09-328-352-6641  
; Sequence 6641, Application US/09328352  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328, 352  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6641  
; LENGTH: 364  
; TYPE: PRF  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6641

Query Match 11.3%; Score 264.5; DB 17; Length 364;  
Best Local Similarity 21.5%; Pred. No. 1,1e-14;  
Matches 87; Conservative 76; Mismatches 152; Indels 89; Gaps 14;

QY 8 IEOLSKPLSDSDSICGVYKLEKSAFRLRNEFNVAQTALRKLSQNPASADERDALOEACLN 67  
Db 5 ISELKRLINDSLCG-----EDYSF---SNEFHEIKKA--RTODDLLDQGDWVAERKQA 54  
QY 68 KWKILSLSYLVEQFSKTRDIELISWFAAOFLDPTLESANSLWMLADISEKHWDLNP 127  
Db 55 DWDFVAVSVSTLLIEKTRDLRLTWIETWHLN-GFEGWYKGITLTHMLNQWODIHP 113  
QY 128 VLPETLKSDDDKEREQADAKVKAFFQVGDSEESSIYAPV--OLPLVGEVTFE-- 183  
Db 114 I-----EDDD-----LDQRIQGLGLNQ-----LPMILKKVPLTNTAPYYNL 152  
QY 184 -----DFOSEAKKEGELSOLKSMLTITVAOERPAIOFKMENAKRCVTO 225  
Db 153 LDYDNFLYHENIRRKQTEVEESQSGPSELQFOAIFNT-----SKTFQYSNYQEFNSV 206  
QY 226 LDRSALVSTCHSLGOSTNFGAKSLITRVENAL-----VHLSGILKPKAKAEKTYEQ 280  
Db 207 LTRNNVVKQITDHLMDGDSFPAIDSAFETIHSTLRKIYKAEAFGGGLASQOAAV-- 264  
QY 281 EVAVSSVSEGLPSHMDTKHIERIPMAEQOQVSOHLHAGNLSLGNLNNMNDLAFHL 340  
Db 265 -----ITTPSMENQVYVQIVSDQPMFOQAQT---HL-----ANREQAMKV 302  
QY 341 LREVSDFROSEPSPISTFLLEKAIKRGVLSLPELLREKMSSEON 384

Db 303 LOEIADYFOANEPHSPVSYMLQKTIKWSQMLHEWMLAQVTKDEH 346

RESULT 3  
US-09-252-991A-25690  
; Sequence 25690, Application US/09252991A  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252, 991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25690  
; LENGTH: 681  
; TYPE: PRF  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25690

Query Match 8.4%; Score 197; DB 16; Length 681;  
Best Local Similarity 20.9%; Pred. No. 5e-08;  
Matches 87; Conservative 79; Mismatches 156; Indels 94; Gaps 16;

QY 1 MPLSKH--QIEOLSKPLSDSDSICGVYKLEKSAFRLRNEFNVAQTALRKLSQNPASADER 58  
Db 318 LPLSGNALSLLEVLPEPIDPGQPCGPPSLRYDPDYDR-----LRELREDDSSLP 365  
QY 59 DALOEACLNK--WKILSDSYEDEFKTTTRDIELISWFAAOFLDPTLESANSLWMLAD 116  
Db 366 TGVQOAKRAKADMAVQIASDLQRRSKDMLAAWIGA-WLQRGSLGGLQALVILAE 424  
QY 117 LSEKHWDLNVLPEVETLKSDDDKEREQADAKVKAFFQVGDSEESSIYAPVLOPL 176  
Db 425 LCERYPEEVHP-----QAQDGDQSWVRPPLDWLIRRYAELH-----TRLPL 466  
QY 177 VG-----EYTFEDFOSEAKERK-----GEISOLKSMLTITVAOERPAIOFKMENAKR----- 221  
Db 467 MCGQFAEITLYAMQRIQROQVAGDSKSAKAALEAAQLOOK-----KIDELALRAEPLVQ 521  
QY 222 -----CVQIDRLSALVSTCHSLGOSTNFGAKSLITRVENALVHLSGILKAP 271  
Db 522 WOKQASLACQOQLOQLREQWCD-RC--LQELAPSCQPLREVIAQWMLALKEF--TAMHP 576  
QY 272 KAEAKTYEQVAVSSVSEGLPSHMDTKHIERIPMAEQOQVSOHLHAGNLSLGNLNN 331  
Db 577 QAPLPEQPPVAEADAEG-----DADGESVVASAPSGPAGP----- 615  
QY 332 MNRDLAHLREVSDFROSEPSPISTFLLEKAIKRGVLSLPELLREKMSSEONGDA 387  
Db 616 TSREDAYROLILLADYLRATEPHSPVPVYLKRAVEWGNKPLSELLAELI---NADS 668

RESULT 4  
US-09-649-996-14  
; Sequence 14, Application US/09649996  
; GENERAL INFORMATION:  
; APPLICANT: Jono, Keith E.  
; TITLE OF INVENTION: KINASE GENES AND USES  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; Suite 4700  
; City: Los Angeles  
; State: California  
; Country: U.S.A.



Db 47 DQEPAP-----VEOTPAIA--ACTDESGGSAARTSPASPECISSRDEAF 88  
QY 339 HLLREVSDFROSEPHSIFLEKKAIRMGYLSPELLEKMMSE 382  
Db 89 ETLISVAREFRTEREPHSISLIEFIVRGRMDSSELLAEELLE 132

RESULT 7  
US-09-091-501-B  
; Sequence 8, Application US/09091501  
; GENERAL INFORMATION:  
; APPLICANT: Tinsley, Jonathon M  
; APPLICANT: Davies, Kay E  
; TITLE OF INVENTION: Ultraphin gene expression  
; FILE REFERENCE: 620-42  
; CURRENT APPLICATION NUMBER: US/09/091,501  
; EARLIER FILING DATE: 1998-07-14  
; EARLIER APPLICATION NUMBER: PCT/GB96/03156  
; EARLIER FILING DATE: 1996-12-19  
; EARLIER APPLICATION NUMBER: GB 9525962.8  
; EARLIER FILING DATE: 1995-12-19  
; EARLIER APPLICATION NUMBER: GB 9615797.9  
; EARLIER FILING DATE: 1996-07-26  
; EARLIER APPLICATION NUMBER: GB 9622174.2  
; EARLIER FILING DATE: 1996-10-24  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 2008  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric  
US-09-091-501-B

Query Match 5.9%; Score 138; DB 14; Length 2008;  
Best Local Similarity 21.2%; Pred. No. 0.08;  
Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19;

QY 1 MPLSKHQIEQLSKPLSDSICGVYIKLEKSAFRLRNEFNVAQTALRKLSONSPSADERDA 60  
Db 420 MELQKQLOQLSSWLA-----LTERIQKMSPEPLGDDLP 455  
QY 61 LQEAELKMKWILSDLSYEQSKTTRDIELI-----SMFVAQFLDITLESANSLFWLA 115  
Db 456 LQK-LLOEHKSLQNDLAEQVKVNSLTHMVYIDENSGESATLLEQLOKLGEMTAVC 514  
QY 116 DLSEKHHDLNPLVPVETLKSDDDKGERQADAKVKAFLQVGDSESSITVAPVQLP 175  
Db 515 RWTERRNRL-----OEISILMOELLEQ 538  
QY 176 LVGEVTFEFDQSAERKGEISQLKSMLTITVAQERFAIOFKMNAKRCVQDLRLSAL-- 232  
Db 539 CLLEAWLKEKEALNKVQTSNFKQKELSVYRLATILKEDEMKR--QTLDOQLSEIGOD 596  
QY 233 -----VSTKCHSLSGOSTNGFAKSLITRVENALVHLSGIKLAPKAEKTVQDEVA 283  
Db 597 VQQLSNPKASKKNNSSEELTORW--DSLVRLEDS-----SNQVTOAVA 640  
QY 284 ESSVSEGEELSHMDTKHIERIPMASEOAO-----TVSOH-LHAGNLSLGNLNNMRDL 336  
Db 641 KLGSQIPQKDLLETIVAVREKGMVKKPKQELPPLTKAEHAKMRSTTELG----- 691  
QY 337 AFHLREVSDFROSEPH-----SPISFLEKKAIRMGYLSPELLEKMMSEONGDA 387  
Db 692 --ENLQELRLDLOQMEVHAELKWLNTLEMLSDKS-----LSLPE--RDKISE-----S 738  
QY 388 LSTFNAAGNLHDQVLLPEVSTPTVIGIESPQTPQAKPSVSDRPSVEHVSQTSPPVDQTS 447  
Db 739 LRTV-----NMTWKNKICREV--PTTLKECIOEPS--SVSQTR-IAAHNVQKVVVLVSS 786  
QY 448 KODQPOSSATSALS 462

Db 787 ASDIPVOSHRTSEIS 801

RESULT 8  
US-09-091-501B-B  
; Sequence 8, Application US/09091501B  
; GENERAL INFORMATION:  
; APPLICANT: Tinsley, Jonathon M  
; APPLICANT: Davies, Kay E  
; TITLE OF INVENTION: Ultraphin gene expression  
; FILE REFERENCE: 620-42  
; CURRENT APPLICATION NUMBER: US/09/091,501B  
; EARLIER FILING DATE: 1998-06-18  
; EARLIER APPLICATION NUMBER: PCT/GB96/03156  
; EARLIER FILING DATE: 1996-12-19  
; EARLIER APPLICATION NUMBER: GB 9525962.8  
; EARLIER FILING DATE: 1995-12-19  
; EARLIER APPLICATION NUMBER: GB 9615797.9  
; EARLIER FILING DATE: 1996-07-26  
; EARLIER APPLICATION NUMBER: GB 9622174.2  
; EARLIER FILING DATE: 1996-10-24  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 2008  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; NAME/KEY: misc\_feature  
; LOCATION: (239) ... (250)  
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric; Xaa = Unknown  
US-09-091-501B-B

Query Match 5.9%; Score 138; DB 14; Length 2008;  
Best Local Similarity 21.2%; Pred. No. 0.08;  
Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19;

QY 1 MPLSKHQIEQLSKPLSDSICGVYIKLEKSAFRLRNEFNVAQTALRKLSONSPSADERDA 60  
Db 420 MELQKQLOQLSSWLA-----LTERIQKMSPEPLGDDLP 455  
QY 61 LQEAELKMKWILSDLSYEQSKTTRDIELI-----SMFVAQFLDITLESANSLFWLA 115  
Db 456 LQK-LLOEHKSLQNDLAEQVKVNSLTHMVYIDENSGESATLLEQLOKLGEMTAVC 514  
QY 116 DLSEKHHDLNPLVPVETLKSDDDKGERQADAKVKAFLQVGDSESSITVAPVQLP 175  
Db 515 RWTERRNRL-----OEISILMOELLEQ 538  
QY 176 LVGEVTFEFDQSAERKGEISQLKSMLTITVAQERFAIOFKMNAKRCVQDLRLSAL-- 232  
Db 539 CLLEAWLKEKEALNKVQTSNFKQKELSVYRLATILKEDEMKR--QTLDOQLSEIGOD 596  
QY 233 -----VSTKCHSLSGOSTNGFAKSLITRVENALVHLSGIKLAPKAEKTVQDEVA 283  
Db 597 VQQLSNPKASKKNNSSEELTORW--DSLVRLEDS-----SNQVTOAVA 640  
QY 284 ESSVSEGEELSHMDTKHIERIPMASEOAO-----TVSOH-LHAGNLSLGNLNNMRDL 336  
Db 641 KLGSQIPQKDLLETIVAVREKGMVKKPKQELPPLTKAEHAKMRSTTELG----- 691  
QY 337 AFHLREVSDFROSEPH-----SPISFLEKKAIRMGYLSPELLEKMMSEONGDA 387  
Db 692 --ENLQELRLDLOQMEVHAELKWLNTLEMLSDKS-----LSLPE--RDKISE-----S 738  
QY 388 LSTFNAAGNLHDQVLLPEVSTPTVIGIESPQTPQAKPSVSDRPSVEHVSQTSPPVDQTS 447  
Db 739 LRTV-----NMTWKNKICREV--PTTLKECIOEPS--SVSQTR-IAAHNVQKVVVLVSS 786  
QY 448 KODQPOSSATSALS 462  
Db 787 ASDIPVOSHRTSEIS 801

```
RESULT 9
US-09-252-991A-17293
; Sequence 17293, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17293
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17293
```

```
Query Match          5.9%; Score 137.5; DB 16; Length 357;
Best Local Similarity 21.6%; Pred. No. 0.0046;
Matches 91; Conservative 67; Mismatches 145; Indels 119; Gaps 18;
```

```
QY 15 LSSDSTICGYTLKLEKSAFRLRNEFVAQTLRLKLSO--NPSADERALQECINKKILS 73
DB 24 VSPDSCGDDL-----ETDAFLLELRINAGOPEROAGAVLPAEPPEPRVR 71
QY 74 DSYEOPSKTRDIELISMFVAQFLDITLESANSLFMDLSEKHNDLNPVLPVET 133
DB 72 ALASELFGR-SKDLRVANLLQSNVLD-GDGLADGILLVRLLDQYMDGYPL----- 124
QY 134 LMSDDKKEKREDAKVAFFQVLQVDSSESSITLAPVQL---PLVGEVTF--FDPOS 187
DB 125 LQADD-----NDPFRIALTGLVAE-----PLDQWVAIPLVSRAGPYNLRA 170
QY 188 AERKGEISOLKSMLTITVAOERPAIOFKMENAKRCVTQDRLSALVSTCHSGSOSTNF 247
DB 171 ALNAGLQFPAS---ETLSPEQIAGAFADADA-----DALATRAALEGAGOEHAL 217
QY 248 GFAKSLITREVENA-----LVHL-----SGIKLAPKAEKTYEQEVAES 285
DB 218 ATESSGVAERVSAQGLDGLPLRQLRLQALQVFDLYGPGAGERLACAGAAADQOGA-A 276
QY 286 SVSEGLPESHMDTKHIERIPMASEQAQTVSQHLHAGNLSLGNLNMNMDLAEFHLLREVS 345
DB 277 PYAAVAAPA-----PRAS-----GEIA-----NREDVLRQDRL 306
QY 346 DYFROSEPHSPISFILEKAIKRCYLSLPELLRMESEQNDALSTTFNAGLNHLDOVLL 405
DB 307 EYIVRHEPSSPVVLLKRAKTLVTADFAEIVRNLIPD-----GISQFETLRG 353
QY 406 PE 407
DB 354 PE 355
```

```
RESULT 10
US-60-173-464-18560
; Sequence 18560, Application US/60173464
; GENERAL INFORMATION:
```

```
; APPLICANT: LI, Peter W. D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; FILE REFERENCE: CLO00173
; CURRENT APPLICATION NUMBER: US/60/173,464
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 18560
```

```
; LENGTH: 5884
; TYPE: PRT
; ORGANISM: Drosophila
US-60-173-464-18560
```

```
Query Match          5.8%; Score 136.5; DB 27; Length 5884;
Best Local Similarity 22.0%; Pred. No. 0.69;
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;
```

```
QY 30 SARPLRNE-FNVAQTA---LRKLS-----QNPASD---ERP---ALQEA---CLNKWK 70
DB 1259 SEFVDKKNDFSMLOETALRSLRTPLOTPDKNVSDLSKRLNVOLOQASHQLLPKLH 1318
QY 71 ILSDSYEOPSKTRDIELISMFVAQFLDITLESANSLFMDLSEKHNDLNPVLPVET 124
DB 1319 ALKSELAPLAAPDKRPI-LEKEVTEVKEFNTMEHWKDVGLYEDYSAK-MNNYTRRLA 1376
QY 125 -----LNPLP--VETLKSDDDKKEREQADAKVKAFFQVLQD--SESSITLAPVQL-P 175
DB 1377 ELQEMANKVAPKNIKIALQSEDLTPREVR---VKVQAFKRLIGDRMKQDLDLADASELAP 1433
QY 176 LVGEVTFPDEQSAERKGEISOLKSMILT----- 203
DB 1434 KEGNIA-----EAKRLKGEITKLOEVLSAIRNVHDQAQVQEDLVNMQFQAGLQOIKPA 1489
QY 204 -----TVAOERPAIOFKMENAKRCVTQDRLSALVSTCHSGS 243
DB 1490 VEOSEVKNVNVSKPISLEBAVAMQNAQOFETQ---COEQDILHG-1SNISHKMLCK 1544
QY 244 STNFGFAKSLITREVEN--ALVHLGSIKLAPKABA-----KTVEQEVA--ESSVSEGE 291
DB 1545 -TN---APELDAMHSRMTAVHENAQASAKLEKLVANKMSFPADAAKLEWDWGQEQDM 1600
QY 292 --LPISHMDTKHIERI-----PMASDAQTV-----SQHLHAGNLSL 326
DB 1601 SRRPAVLNTPPHIDKLEKELVKLSFNNNEISQOQAKLVTLQGNADOQISLHAPGGAALKD 1660
QY 327 -----GNLNMNRDLAFHLLREVSDYF--ROSEPHSPISFILEKAIKRCYLSLPELLRE 378
DB 1661 RVNOMKKGLQKLEATRGH-INEVSDIITSRODPNAKLVE-----SNW-----MDQLRN 1709
QY 379 MMS---EQNGDALST---TFNAGLNHLDO-----VLLPEVSTPYGIESPOTPOAKPSV 427
DB 1710 QVTQVEINERETSLHVIHALLOHADKKRPFNMAYIDVKKDALG---ATPESNML 1765
QY 428 SDP-----RSVEEHVSQTSV-----DTQS-----KODKPOSSATSAL 461
DB 1766 NDAYTALVYVQNLFTNMLOKKAALBKWTELLGKKNDTESHLNYLKHQLDKREGPAAEEL 1825
QY 462 S 462
DB 1826 S 1826
```

```
RESULT 11
US-60-167-217-22879
; Sequence 22879, Application US/60167217
; GENERAL INFORMATION:
```

```
; APPLICANT: LI, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CLO00152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 22879
; LENGTH: 7118
; TYPE: PRT
; ORGANISM: Drosophila
US-60-167-217-22879
```

Query Match 5.8%; Score 136.5; DB 27; Length 7118;  
Best Local Similarity 22.0%; Pred. No. 0.96;  
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;

```
QY 30 SAFRPLRNE-FNVAQTA---LRKLS-----QNSAD---ERD---ALQEA---CLNKKW 70
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1259 SEFDQKNDJFSMLQTAETELRSLTPLOTDPKKNVSQDLKSKRDNLVNOLOQASHOLLPKLH 1318
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 71 ILSDSLYEQFSKTTTRDIELISWFVAQAQFLDPTLESANSLMWLADISEKHMDH----- 124
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1319 ALKSELAPLAPDKRP1-LEKEVTEVEKMFENMEHKDRGYLEDYSAK-WNNYKTRLA 1376
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 125 -----LNPVLPR-VETLKSDDKEREQADAKYAFQVLGD-SESSIIYAPVLQ-LP 175
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1377 ELQEMAKVAPKNIEALQSEDLTPPEERV---VKVQAFKRITIGDMKOLDLLAADASELAP 1433
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 176 LVGEVTFEFDQSAERKEISQKSMLT----- 203
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1434 KEGNIA---EAKRLKEITRLQEVLSAINRNVDHOAAVOEDLVNMQQFQAGLQOIKPA 1489
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 204 -----TVAQERFAIOFKMENAKRCYTQDLRLSALVSTKCHSLSGQ 243
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1490 VEOSEVAVNNVNSKPISELEVAVMAQNAQOPEIQ---CQEQDLKLING-1SNISHKMLCK 1544
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 244 STNGFPAKSLITRYEN-ALVHLSGILAPKAE-----KTVEQEA-ESSVSEGE--- 291
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1545 -TN---APDELDAMHSMTAVHENAQAASAKLEKLVANMKSFDADAQKLEDMVWGQEQOM 1600
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 292 --LPSHDTHIERI-----PMASEQAQTV-----SOHLHAGNLSTL- 326
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1601 SRRPAVINTPHIDLEKELVKLKSFNNEISQQAQKLVTLGONADOISLHLAPBEGAALKD 1660
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 327 -----GNLNNMNDLAFHLREVSDF--ROSEPHSPISFLKAIKRWGYLSLPELIRE 378
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1661 RVNOMKGLKLOKLSATRGH-INEVSDAISRQDENAKLVN-----SNW-----MEQLRN 1709
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 379 MMS---EQNGDALST---IFNAAGLNHLQD-----VLLPEVSTPTVGIESQTPQAKPSV 427
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1710 QVTVQEEINPERVETSLVHIALQEHADKKPSFNALYDEVKQALG---ATPEESNAL 1765
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 428 SDP-----RSVEEHVSQTSVP-----PTQS-----KODKPOSSATSAL 461
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1766 NDATVATLVNQTLETNMLQKKALEKTELLGKNKNDTESHLNKLHQDLKPBPAPAEEL 1825
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 462 S 462
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1826 S 1826
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 12
US-60-173-464-20674
; Sequence 20674, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; FILE REFERENCE: C1000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; NUMBER OF SEQ. ID NOS: 1999-12-29
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 20674
; LENGTH: 7118
; TYPE: PRT
; ORGANISM: Drosophila
US-60-173-464-20674

Query Match 5.8%; Score 136.5; DB 27; Length 7118;
Best Local Similarity 22.0%; Pred. No. 0.96;
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;
```

```
Db 1259 SEFDQKNDJFSMLQTAETELRSLTPLOTDPKKNVSQDLKSKRDNLVNOLOQASHOLLPKLH 1318
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 71 ILSDSLYEQFSKTTTRDIELISWFVAQAQFLDPTLESANSLMWLADISEKHMDH----- 124
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1319 ALKSELAPLAPDKRP1-LEKEVTEVEKMFENMEHKDRGYLEDYSAK-WNNYKTRLA 1376
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 125 -----LNPVLPR-VETLKSDDKEREQADAKYAFQVLGD-SESSIIYAPVLQ-LP 175
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1377 ELQEMAKVAPKNIEALQSEDLTPPEERV---VKVQAFKRITIGDMKOLDLLAADASELAP 1433
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 176 LVGEVTFEFDQSAERKEISQKSMLT----- 203
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1434 KEGNIA---EAKRLKEITRLQEVLSAINRNVDHOAAVOEDLVNMQQFQAGLQOIKPA 1489
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 204 -----TVAQERFAIOFKMENAKRCYTQDLRLSALVSTKCHSLSGQ 243
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1490 VEOSEVAVNNVNSKPISELEVAVMAQNAQOPEIQ---CQEQDLKLING-1SNISHKMLCK 1544
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 244 STNGFPAKSLITRYEN-ALVHLSGILAPKAE-----KTVEQEA-ESSVSEGE--- 291
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1545 -TN---APDELDAMHSMTAVHENAQAASAKLEKLVANMKSFDADAQKLEDMVWGQEQOM 1600
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 292 --LPSHDTHIERI-----PMASEQAQTV-----SOHLHAGNLSTL- 326
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1601 SRRPAVINTPHIDLEKELVKLKSFNNEISQQAQKLVTLGONADOISLHLAPBEGAALKD 1660
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 327 -----GNLNNMNDLAFHLREVSDF--ROSEPHSPISFLKAIKRWGYLSLPELIRE 378
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1661 RVNOMKGLKLOKLSATRGH-INEVSDAISRQDENAKLVN-----SNW-----MEQLRN 1709
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 379 MMS---EQNGDALST---IFNAAGLNHLQD-----VLLPEVSTPTVGIESQTPQAKPSV 427
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1710 QVTVQEEINPERVETSLVHIALQEHADKKPSFNALYDEVKQALG---ATPEESNAL 1765
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 428 SDP-----RSVEEHVSQTSVP-----PTQS-----KODKPOSSATSAL 461
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1766 NDATVATLVNQTLETNMLQKKALEKTELLGKNKNDTESHLNKLHQDLKPBPAPAEEL 1825
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 462 S 462
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1826 S 1826
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 13
US-60-191-637-39835
; Sequence 39835, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; FILE REFERENCE: C1000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; NUMBER OF SEQ. ID NOS: 42660
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 39835
; LENGTH: 7182
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-637-39835

Query Match 5.8%; Score 136.5; DB 27; Length 7182;
Best Local Similarity 22.0%; Pred. No. 0.97;
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;
```

```
Db 1323 ALKSELAPLAADPKRPI-LEKEVTEVEKMFENTMEHYKDRVGYLEDYSAK-WNNYKTRLA 1380
Qy 125 -----LNPLP--VETLKSDDDKGEREQADAKVKAFFOLVGD-SESSITLAPVLQ-L-P 175
Db 1381 ELQEMANKVAPKNIELOSEDLTPREVR--VKVQAFKRILGDRMKQDLDAADASELAP 1437
Qy 176 LVGEVTFEFDQSAERKEISQLKSMILT----- 203
Db 1438 KEGNIA-----EAKRLKEITKLOEVLISAIRNVDHOQAQVOEDLVNMQOFOAGLQIKRA 1493
Qy 204 -----TVAQERFATQFKMENAKRCVTQDLRLSALVSTKCHSLGSO 243
Db 1494 VQOSEVKNVNVSKPISLEEAVALMOQAOFFETQ-----COEOLDKLHG-ISMNISHKMLCK 1548
Qy 244 STNFGAKSLTRVEN--ALVHLSGIKLPKAE-----KTYQGEVA--ESSVSEGE----- 291
Db 1549 -TN---APDELDAMHSRMTAVHENAQAASAKLEKLVANMKSFDADAKLEDWVGQGEQOM 1604
Qy 292 --LPSHMDTKHIERI-----PMASEQAQTV-----SOHLHAGNLSEL-- 326
Db 1605 SRPVALNTPHIDKLEKELVTKLSFNNEISQOQAKLVITGQANADQISLHAPGGAALAD 1664
Qy 327 -----GNINNMNRDLAFHLLREVSDYF--ROSEPHSPISFLEKAIIRNGYLSLPELRE 378
Db 1665 RVNQMGKLOKLTSEATRGH-INEVSDAISRODFNAKLVNF-----SNM-----MEQLRN 1713
Qy 379 MMS-----EONGDALST---IFNAAGLNHLDO-----VLPEVSTPTVGIESPOTPOAKPSV 427
Db 1714 QYQVEIENPERVETSLAHVIALHLOEHADKPSFNAIYDEVKOLALG---ATPEESNML 1769
Qy 428 SDP-----RSVEEHVSQTSVPV-----DTOS-----KODOKPOSSATSAL 461
Db 1770 NDAVYALVNVYQNLFTNMLQKKALEKMTLLGKWKNDTSHLNYLKHQDLKPGPAAEEL 1829
Qy 462 S 462
Db 1830 S 1830

RESULT 14
US-60-191-681-30862
; Sequence 30862, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF.
; FILE REFERENCE: C1000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30862
; LENGTH: 7182
; TYPE: PRP
; ORGANISM: DROSOPHILA
US-60-191-681-30862

Query Match 5.8%; Score 136.5; DB 27; Length 7182;
Best Local Similarity 22.0%; Pred. No. 0.97;
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;
```

```
Qy 176 LVGEVTFEFDQSAERKEISQLKSMILT----- 203
Db 1438 KEGNIA-----EAKRLKEITKLOEVLISAIRNVDHOQAQVOEDLVNMQOFOAGLQIKRA 1493
Qy 204 -----TVAQERFATQFKMENAKRCVTQDLRLSALVSTKCHSLGSO 243
Db 1494 VQOSEVKNVNVSKPISLEEAVALMOQAOFFETQ-----COEOLDKLHG-ISMNISHKMLCK 1548
Qy 244 STNFGAKSLTRVEN--ALVHLSGIKLPKAE-----KTYQGEVA--ESSVSEGE----- 291
Db 1549 -TN---APDELDAMHSRMTAVHENAQAASAKLEKLVANMKSFDADAKLEDWVGQGEQOM 1604
Qy 292 --LPSHMDTKHIERI-----PMASEQAQTV-----SOHLHAGNLSEL-- 326
Db 1605 SRPVALNTPHIDKLEKELVTKLSFNNEISQOQAKLVITGQANADQISLHAPGGAALAD 1664
Qy 327 -----GNINNMNRDLAFHLLREVSDYF--ROSEPHSPISFLEKAIIRNGYLSLPELRE 378
Db 1665 RVNQMGKLOKLTSEATRGH-INEVSDAISRODFNAKLVNF-----SNM-----MEQLRN 1713
Qy 379 MMS-----EONGDALST---IFNAAGLNHLDO-----VLPEVSTPTVGIESPOTPOAKPSV 427
Db 1714 QYQVEIENPERVETSLAHVIALHLOEHADKPSFNAIYDEVKOLALG---ATPEESNML 1769
Qy 428 SDP-----RSVEEHVSQTSVPV-----DTOS-----KODOKPOSSATSAL 461
Db 1770 NDAVYALVNVYQNLFTNMLQKKALEKMTLLGKWKNDTSHLNYLKHQDLKPGPAAEEL 1829
Qy 462 S 462
Db 1830 S 1830
```

```
RESULT 15
US-09-614-150-40200
; Sequence 40200, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: C1000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40200
; LENGTH: 7201
; TYPE: PRP
; ORGANISM: DROSOPHILA
US-09-614-150-40200

Query Match 5.8%; Score 136.5; DB 20; Length 7201;
Best Local Similarity 22.0%; Pred. No. 0.98;
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;
```

```

QY 30 SAERPLANE-FNVAQTA---LRKLS-----QNP SAD---ERD---ALQEA---CLNKWK 70
Db 1263 SEFYDQKNDIFSMLOTAETELRSILTPLOTDPKNTVSQDLKSKRDJNVLOQAASHQLLPRLH 1322
QY 71 ILSDSLTYEQFSKTRDIELISMFAAOFLDITLESANSLLEMLADSEKHWDH----- 124
Db 1323 ALKSELAPLAPDKRPIT-LEKEVTEVEKMFENIMEHYKDRVGLLEDYSAK-WNNYKTRLA 1380
QY 125 -----LNPVLP--VETLKSDDDKGKEREQADAKYKAFVOLVD--SEESSIUYAPVLOL-P 175
Db 1381 ELQEMANKVAPKNIEALQSELTPPEERV---VKYQAFKRITIGDRMKQDLADASELAP 1437
QY 176 LVGEVTFPFDSAEKKEBISQKSMLT----- 203
Db 1438 KEGNIA---EAKRLKEITFKLEVLISAINRNDHOQAQVEDLVNMQOFQAGLOITKPA 1493
QY 204 -----TVAOERFALQFMENAKRCVTQDLRLSALVSTKCHSLGSO 243
Db 1494 VEQSEVKNVNVSPKPILEEAVAMQONAOQETQ---COBOLDKLHG-ISHKMLCK 1548
QY 244 STNFGFAKSLTTRVEN--ALVHLSGIKLAPKAE-----KTVEOEVA--ESSVSEGE--- 291
Db 1549 -TN---APDELDAMHSRMTAVHENAKQASAKLEKLVANWKSFPDADAKLEDMVGQGEQOM 1604
QY 292 --LPSHMDTKHIERI-----PMASEQAQTV-----SOHLHAGNISEL-- 326
Db 1605 SRBPVAVLNTPHIDKLEKELVYKLSFNNEISQOQAKLVTLGONADQISLHAPEGAAALKD 1664
QY 327 -----GNLNMNRDLEFHLIREVSDYF--ROSEPHSIFISFLEKAIKRWGLSLPELIRE 378
Db 1665 RVNOMKGRLOKLEATRGH--INEVSDAIIISRODENAKLVNF-----SNW-----MEOLRN 1713
QY 379 MMS---FONGDALST---IFNAGLNLHDQ-----VLLPEVSTPTVGIESPOTPOAKPSV 427
Db 1714 QVTQVEELINPERVETSLHVIHALQEHADKKPSFNATYDEYKQLALG---ATPESNAL 1769
QY 428 SDP-----RSVEEHVSQTSVP-----DTQS-----KODOKPOSSATSAL 461
Db 1770 NDATYALVNVNQNLETNMLQKKALEKTELLGWNKNDTESHLYLTKHQLDKPEGPAAEBL 1829
QY 462 S 462
Db 1830 S 1830

```

Search completed: March 2, 2003, 05:10:57  
 Job time : 283 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 04:11:48 ; Search time 42 Seconds  
(without alignments)  
1027.381 Million cell updates/sec

Title: US-09-915-706A-2

Perfect score: 2343

Sequence: 1 MPLSKHQIEQSLKPLSDSI.....DNQSKQDQKPOSSATSLSW 463

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 494224 seqs, 93196565 residues

Total number of hits satisfying chosen parameters: 494224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PC1\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID                   | Description        |
|------------|-------|-------------|--------|----------------------|--------------------|
| 1          | 143   | 6.1         | 884    | US-10-179-766-10     | Sequence 10, Appl  |
| 2          | 131   | 5.6         | 1329   | PC1-US03-01772-29    | Sequence 29, Appl  |
| 3          | 129.5 | 5.5         | 860    | US-10-072-012-838    | Sequence 838, Appl |
| 4          | 126.5 | 5.4         | 752    | US-60-423-586-154    | Sequence 154, Appl |
| 5          | 126.5 | 5.4         | 752    | US-60-427-194-154    | Sequence 154, Appl |
| 6          | 126.5 | 5.4         | 785    | US-10-309-851-24     | Sequence 24, Appl  |
| 7          | 126.5 | 5.4         | 796    | US-09-724-676-58406  | Sequence 58406, A  |
| 8          | 126.5 | 5.4         | 796    | US-09-724-676-58406  | Sequence 58406, A  |
| 9          | 126.5 | 5.4         | 810    | US-09-724-676-58407  | Sequence 58407, A  |
| 10         | 126.5 | 5.4         | 810    | US-09-724-676-58407  | Sequence 58407, A  |
| 11         | 126.5 | 5.4         | 810    | US-09-724-676A-58408 | Sequence 58408, A  |
| 12         | 126.5 | 5.4         | 810    | US-09-724-676A-58408 | Sequence 58408, A  |
| 13         | 126.5 | 5.4         | 1133   | US-10-309-851-14     | Sequence 14, Appl  |
| 14         | 126.5 | 5.4         | 1133   | US-10-309-851-16     | Sequence 16, Appl  |
| 15         | 126.5 | 5.4         | 1135   | US-10-309-851-12     | Sequence 12, Appl  |
| 16         | 121.5 | 5.2         | 2642   | US-09-724-676-55851  | Sequence 55851, A  |
| 17         | 121.5 | 5.2         | 2642   | US-09-724-676A-55851 | Sequence 55851, A  |
| 18         | 121.5 | 5.2         | 2722   | US-09-724-676-55868  | Sequence 55868, A  |
| 19         | 121.5 | 5.2         | 2722   | US-09-724-676A-55868 | Sequence 55868, A  |
| 20         | 121.5 | 5.2         | 2915   | US-09-724-676-55849  | Sequence 55849, A  |
| 21         | 121.5 | 5.2         | 2915   | US-09-724-676-55850  | Sequence 55850, A  |
| 22         | 121.5 | 5.2         | 2915   | US-09-724-676A-55849 | Sequence 55849, A  |
| 23         | 121.5 | 5.2         | 2915   | US-09-724-676A-55850 | Sequence 55850, A  |
| 24         | 121.5 | 5.2         | 2929   | US-09-724-676-55846  | Sequence 55846, A  |
| 25         | 121.5 | 5.2         | 2929   | US-09-724-676A-55846 | Sequence 55846, A  |
| 26         | 121.5 | 5.2         | 2940   | US-09-724-676-55847  | Sequence 55847, A  |

|    |       |     |      |                      |                   |
|----|-------|-----|------|----------------------|-------------------|
| 27 | 121.5 | 5.2 | 2940 | US-09-724-676-55848  | Sequence 55848, A |
| 28 | 121.5 | 5.2 | 2940 | US-09-724-676A-55847 | Sequence 55847, A |
| 29 | 121.5 | 5.2 | 2940 | US-09-724-676A-55848 | Sequence 55848, A |
| 30 | 121.5 | 5.2 | 2995 | US-09-724-676-55867  | Sequence 55867, A |
| 31 | 121.5 | 5.2 | 2995 | US-09-724-676-55867  | Sequence 55867, A |
| 32 | 121.5 | 5.2 | 2995 | US-09-724-676A-55866 | Sequence 55866, A |
| 33 | 121.5 | 5.2 | 2995 | US-09-724-676A-55867 | Sequence 55867, A |
| 34 | 121.5 | 5.2 | 3009 | US-09-724-676-55863  | Sequence 55863, A |
| 35 | 121.5 | 5.2 | 3009 | US-09-724-676A-55863 | Sequence 55863, A |
| 36 | 121.5 | 5.2 | 3020 | US-09-724-676-55864  | Sequence 55864, A |
| 37 | 121.5 | 5.2 | 3020 | US-09-724-676-55865  | Sequence 55865, A |
| 38 | 121.5 | 5.2 | 3020 | US-09-724-676A-55864 | Sequence 55864, A |
| 39 | 121.5 | 5.2 | 3020 | US-09-724-676A-55865 | Sequence 55865, A |
| 40 | 121.5 | 5.2 | 3259 | US-10-229-059-7      | Sequence 7, Appl  |
| 41 | 118.5 | 5.1 | 1896 | US-09-950-084-6335   | Sequence 6335, Ap |
| 42 | 118   | 5.0 | 674  | US-09-724-676-51054  | Sequence 51054, A |
| 43 | 118   | 5.0 | 674  | US-09-724-676-51055  | Sequence 51055, A |
| 44 | 118   | 5.0 | 674  | US-09-724-676A-51054 | Sequence 51054, A |
| 45 | 118   | 5.0 | 674  | US-09-724-676A-51055 | Sequence 51055, A |

## ALIGNMENTS

```
RESULT 1
US-10-179-766-10
: Sequence 10, Application US/10179766
: GENERAL INFORMATION:
: APPLICANT: Kenyon, Cynthia
: APPLICANT: Apfeld, Javier
: APPLICANT: Dillan, Andrew
: APPLICANT: Garigan, Delia
: APPLICANT: Hsu, Ao-Lin A.
: APPLICANT: Lehrer-Graiser, Josh
: APPLICANT: Murphy, Coleen
: TITLE OF INVENTION: The Regents of the University of California
: FILE REFERENCE: 023070-119310US
: CURRENT APPLICATION NUMBER: US/10/179,766
: CURRENT FILING DATE: 2002-06-24
: PRIOR APPLICATION NUMBER: US 60/300,577
: PRIOR FILING DATE: 2001-06-22
: PRIOR APPLICATION NUMBER: US 60/301,052
: PRIOR FILING DATE: 2001-06-25
: PRIOR APPLICATION NUMBER: US 60/373,975
: PRIOR FILING DATE: 2002-04-18
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 10
: LENGTH: 884
: TYPE: PRT
: ORGANISM: Caenorhabditis elegans
: FEATURE:
: OTHER INFORMATION: 11w-4 (long-lived worm protein), gene F45H10.4
US-10-179-766-10

Query Match      6.1% Score 143; DB 6; Length 884;
Best Local Similarity 20.7%: Pred. No. 0.00074;
Matches 103; Conservative 99; Mismatches 165; Indels 130; Gaps 25;

OY 27 LEKSAAPRPRNEFNVAQTLRLKLNQPSADERDALQDACLNKV---KILSDSLYQFQSK 82
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 218 LRPDATTEARIRRLRLATLSTATLSOR-----REFRLKYSQIRQVVAQNLFFQOOR 268
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 83 TTRDIELISWYAAQFLDITLESANSLFWLADLSKEKIMHDLPVLPVETLKSDDQCK 142
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 269 LRPRVLDLVTETSSKLAVEAPEETTTQDETITDASE-----VTTTKAVEATE 317
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 143 E-REQADAVKAFQVLQGSSESSILYAVLQ-----LPLVGEVTFPD-PQ 186
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 318 EYTEATEATEA---PVATTKESSEMHVNTIRIMIRASAKDLKTYVTLISEKFESELE 374
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 187 SAERK-----GEISQLKSLMLTTTVAQ-----ERFAIQFKENAKRCVTLQ 226
```

Db 375 LADKKTLLTSKDEKISSMAKLKDLINEALSEKESGEIEQAMEKFEKPEKSELYAME 434  
QY 227 DRLSALVSTCHSLGSGSTNGEFAKSLTR----VENALVHLSGIKLAPKAFAKTVQOE 281  
Db 435 DKDIPRAVFTISDSIKHKAELAKLAHTIVSRNVYAENA-----IEKEVEYPRAEKKEVED 491  
QY 282 ---VAESSVE--GELPSHMDTKHIERIPMASEQOATVSQHLHAGLSLGNLNN--N 333  
Db 492 VKAAEKKKEKKRGLP--MKIEKLEK-PVDFK-----SENHELKVLVD 534  
QY 334 RDLAFHLLREVSDFROSEPHSPISFLEK---AIRMGYLSLPPELLREMMSEQ---NGDA 387  
Db 535 KERLLVSESLKNTAEETKPRVE-SFKSEETVVID----DMPALEKESAEKKEETGEP 589  
QY 388 LSTIFNAGLNHLDOVLLPEVSTPVGIESPQPAKPSVSDPSVEEHVSQTSVPDQO 446  
Db 590 TTT-----EAAVETTEASETPPEAKPELLS--NLBDVLTITTP-ETET 630  
QY 447 ---SKODOKPOSSATSA 460  
Db 631 IEGSGEREPTTSAPAA 647

## RESULT 2

PCT-US03-01772-29  
; Sequence 29, Application PC/TUS0301772  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.; YUE, Henry;  
; APPLICANT: GRIFFIN, Jennifer A.; RICHARDSON, Thomas W.;  
; APPLICANT: TANG, Y. TOM; THANGAVELU, Kavitha;  
; APPLICANT: FORSYTHE, Ian J.; BECHA, Shanya D.;  
; APPLICANT: CHAWLA, Narinder K.; HAPALIA, April J.A.;  
; APPLICANT: SWARNAKAR, Anita; MARQUIS, Joseph P.;  
; APPLICANT: GORVAD, Ann E.; BAUGHN, Mariah R.;  
; APPLICANT: LU, Dyung Aina M.; ARVIZU, Chandra S.;  
; APPLICANT: KABLE, Amy E.; LEE, Soo Yeun;  
; APPLICANT: RAMKUMAR, Jayalaxmi; JIANG, Xin;  
; APPLICANT: JACKSON, Alan A.; KHARE, Reena;  
; APPLICANT: ELLIOTT, Vicki S.; BULLOCH, Sean;  
; APPLICANT: XU, Yuming; LEE, Sally;  
; APPLICANT: LEHR-MASON, Patricia M.  
; TITLE OF INVENTION: STRUCTURAL AND CYTOSKELETON-ASSOCIATED PROTEINS  
; FILE REFERENCE: PF-1326 PCT  
; CURRENT APPLICATION NUMBER: PCT/US03/01772  
; CURRENT FILING DATE: 2003-01-16  
; PRIOR APPLICATION NUMBER: US 60/350,702  
; PRIOR FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: US 60/351,715  
; PRIOR FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: US 60/357,402  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: US 60/379,880  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: US 60/381,599  
; PRIOR FILING DATE: 2002-05-17  
; PRIOR APPLICATION NUMBER: US 60/387,270  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: US 60/397,125  
; PRIOR FILING DATE: 2002-07-19  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: PERL Program  
; SEQ ID NO 29  
; LENGTH: 1329  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 7512382CD1  
PCT-US03-01772-29

Query Match 5.6%; Score 131; DB 1; Length 1329;  
Best Local Similarity 22.0%; Pred. No. 0.016;

Matches 108; Conservative 69; Mismatches 182; Indels 132; Gaps 23;  
QY 3 LSKHQIEOLSKPLSDSICGVYTLKESAPRPRNEFNVAQTALRKLSQNSPAD---RD 59  
Db 907 LKAHQVQVAGHNLKEASSASQFELE-IVLKEKENELKRLKLEAMLKREESLQTLQLOD 965  
QY 60 ALQFACLKWKI--LSDSLYQFSKTTIDIELIS-----WFAAQPLDITLLE 105  
Db 966 VQDENKLFKSOIQLOKQOQOASSFPPEHLLKLVISEREKEISGLMNE-----LDSLKD 1020  
QY 106 SAANSLEMLADLSKEDHDLNPVLPVETLQSD--DDGKREDO-----ADAKYKAF 155  
Db 1021 AVEHQKKNNLDREKNEAMEALASTKMLQDYVNTKSKERQOQVEAVELEAKVLEKLF 1080  
QY 156 QLVQDSEESSIIYAPVQLPLVGEVTPFDQSAERKGEISQLKSMLTITVAQERFALQFK 215  
Db 1081 PKV--SVPSNLSYGEWL-----HGEKKAK-----EGMAGTSGSEEVKYLEHK 1121  
QY 216 MENAKRCVQGLDRLSALVSTKCHSLGSGSTNGEFAKSLTRVENALVHLSGIKLAPKAE 275  
Db 1122 LKEA-----DEWHITLQLECEKY-----KSLAETEGIIQKLO----- 1154  
QY 276 KTYEQEVAESSYSRGELPSHMDTKHIERIPMASEQOATVSQHLHAGLSLGNLNNMNRD 335  
Db 1155 RSVEQEEKNKKVAVDE--SHKTIKQMOSSFTSSBQ-----ELERLRSENKD 1198  
QY 336 LAFHLLREVSDFROSEPHSPISFLEKA-----IRMGYLSLPPELLREMMSEON- 384  
Db 1199 I-----ENLRREKH--LEMELEKALMERSTYTVTEVRELKAQNLNLTUTLRPQNE 1247  
QY 385 ----GDALSTIFNAGLNHLDOVLLPEVSTPVGIE-SPOTPAKPSVSDPSVEEHVS 438  
Db 1248 ROKVAGD----LHKAQOQLSLQSKIVKAGDTTVIENSQVSBETESSEKETSQSVS--LN 1301  
QY 439 QTSVPDQOSKO 449  
Db 1302 QTV---TQLOQ 1309

## RESULT 3

US-10-072-012-838  
; Sequence 838, Application US/10072012  
; GENERAL INFORMATION:  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Spletz, Kimberly  
; APPLICANT: Zethusen, Bryan  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Li, Li  
; APPLICANT: Gangoli, Bsha  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Anderson, David W.  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Taupier Jr, Raymond J.  
; APPLICANT: Gusev, Vladimiro Y.  
; APPLICANT: Coleman, Steven D.  
; APPLICANT: Wolenc, Adam R.  
; APPLICANT: Pena, Carol E. A  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Grose, William M.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Riegler, Daniel K.  
; APPLICANT: Burgess, Catherine E.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-258  
; CURRENT APPLICATION NUMBER: US/10/072,012  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 60/265,102  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 60/265,514

;; PRIOR FILING DATE: 2001-01-31  
;; PRIOR APPLICATION NUMBER: 60/265,517  
;; PRIOR FILING DATE: 2001-01-31  
;; PRIOR APPLICATION NUMBER: 60/265,412  
;; PRIOR FILING DATE: 2001-01-31  
;; PRIOR APPLICATION NUMBER: 60/265,395  
;; PRIOR FILING DATE: 2001-01-31  
;; PRIOR APPLICATION NUMBER: 60/266,406  
;; PRIOR FILING DATE: 2001-02-02  
;; PRIOR APPLICATION NUMBER: 60/266,767  
;; PRIOR FILING DATE: 2001-02-05  
;; PRIOR APPLICATION NUMBER: 60/267,057  
;; PRIOR FILING DATE: 2001-02-07  
;; PRIOR APPLICATION NUMBER: 60/266,975  
;; PRIOR FILING DATE: 2001-02-07  
;; PRIOR APPLICATION NUMBER: 60/267,459  
;; PRIOR FILING DATE: 2001-02-08  
;; Remaining Prior Application data removed - See file Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 1391  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO: 838  
;; LENGTH: 860  
;; TYPE: prt  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Myosin tail  
US-10-072-012-838

Query Match  
Best Local Similarity 5.5%; Score 129.5; DB 6; Length 860;  
Matches 98; Conservative 87; Mismatches 208; Indels 137; Gaps 17;

QY 7 QTEQLSKPLSD-----DSTCGVYLKLESAFRLPNEFVNAQTALR 47  
DB 139 QTEQLSKPLSD-----DSTCGVYLKLESAFRLPNEFVNAQTALR 197  
QY 48 KLSQNSADERDLOACLNKWKILSDSYEQFSKTRDIE---LISFVAQFLDPT 103  
DB 198 ELQRO-----LNDLTSQKSRLOSENSDLTRQLEEAQVSNLSKQLESG 244  
QY 104 LESANSLEWLADLSKHHDLNPLVPTLKSDDDKGREGQADAKVKAFFOLVGDSEB 163  
DB 245 LEAKKSL-----EESREANLQALROLRHLHDLSLREOLEESEAFALE---RQ 293  
QY 164 SLLYAPVLOPLVGEVTEFPOSAERKEISQLKSMLTYYAQ----- 207  
DB 294 LSKANAELQOMR-----SKFESEGLRAEELKLLKKNOKISELEEAANAKCDSL 348  
QY 208 -----ERFAIOFKMENAKRCVTQL-----DRLSALVSTKCHSLG 241  
DB 349 KTKSRLOSELEDLOLEERANAAASELEKKKFNFKILAEKRRKVDLQALDPTAGREAR 408  
QY 242 SOSSTNGFAKSLITREYENLVHLSGIKLAPKAETVEQVAE--SSVSEGLPSHMDTK 299  
DB 409 NLSTE-----LFLKNELEELKDQVEALRRENKNLQDEIHDLTQOLGEGCHVHELEK 461  
QY 300 HIERIPMASEQOQVTSQHLAG--NLSELG-----NLNMNRLAHLRLREVSDFYRQSE 352  
DB 462 ARRLAEDELOALEEAELEESVYLRQVELSOIRSEIERLKEEKEEFNTRK 521  
QY 353 PHSPIFLLEKAIKMGYLSLPELRPMSEQNDALSTFNAGLHNLQOVLLPEVSTPT 412  
DB 522 NH-----QRAIE-----SLQATLEAETGKAKA-----SRKKKLEGGDINELE 559  
QY 413 VGIESPOTQAKPSVSDPRSVEEHVSQTSPTVDTQSKODKPOSSATSALS 462  
DB 560 IALDHAN---KANAEOAKNVKKYQOOVKELQTOVEEQARARADAREQOLA 605

RESULT 4  
US-60-423-586-154  
; Sequence 154, Application US/60423586  
; GENERAL INFORMATION:

;; APPLICANT: American Home Products Inc.  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING AND TREATING SCHIZOPHR  
;; FILE REFERENCE: AM101288L  
;; CURRENT APPLICATION NUMBER: US/60/423,586  
;; CURRENT FILING DATE: 2002-11-05  
;; NUMBER OF SEQ ID NOS: 234  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO: 154  
;; LENGTH: 752  
;; TYPE: prt  
;; ORGANISM: Homo sapiens  
US-60-423-586-154

Query Match  
Best Local Similarity 21.9%; Score 126.5; DB 7; Length 752;  
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KHEQLSKPLS-----DSSI-----CGVYLKLESAFRLPNEFVNAQTALR 48  
DB 160 KLEVEKLSKRLMALEKLEDAFNKSKQECYSKCN--LEKERMTTKQLQSOLESLKVRKE 217  
QY 49 LSONPSADERD--ALOACLNKWKILS-----DSLYEQFSKTRDIEILISFVAQFL 99  
DB 218 LEAIESRLKTEFTLTKED--LTKLTLTVMFVDEKRTMSKLTEDKQLAASSQLQVEON 276  
QY 100 LDT-----LESANSLEWLADLSKHHDLNPLVPTLKSDDDKGREGQADAKVKA 154  
DB 277 KYTTYTEKLIETKTRALKSKTDVEEMKY-----SVTKEDDKLKNKLA-- 319  
QY 155 FOLVGDSEESSILYAPVLOPLVGEVTEFPOSAER---KGEISQLKSMLTYYAQERA 211  
DB 320 -----BEKGNDLSRVNML-----KNRLOSLEIENDFLKNLNQSGSTALHGENNK 370  
QY 212 IOFKMENAKRCVTQDLRLSAL-----VSTKCHLSQOSTN--FGAKSLITREYENLVH 264  
DB 371 IKELSOEVRLLKLLKDMKAIEDDLKTEDEYETTLERRVYANERDKAOFLSKELEHVKMEL 430  
QY 265 SGIKLAPKAA-----KTVEQVAVSSVSEGLPSHMDTKIERIP--MASEQOQVTS 315  
DB 431 AKKLEKETSHQWLFKRLQEEAKS-----GHLSREVDALK--EKIHETMATEDLICHL 485  
QY 316 OHLNACNLSELGNLNMNNDLAFLRLREVSDFYRQSEPHSPISFLLEKAIKMGYLSLPEL 375  
DB 486 OGDHSVCKKKLNQOENRNDLG---REIENLTKELERYRHS---KSLR-----PSL 531  
QY 376 LREPMSEQNDALSTFNAGLHNLQOVLLPEVSTPTVGIESPOTQAKPSVSDPRSV 435  
DB 532 -----NGRRISD-----POVFSKEVQTEAVNDEPDYKSLIP-----LER 566  
QY 436 HVSQTSPTVDTQSKODKPOSSATSALS 463  
DB 567 AVINGOLYEBSENQDEDPNDEC--SVLSF 593

RESULT 5  
US-60-427-194-154  
; Sequence 154, Application US/60427194  
; GENERAL INFORMATION:  
; APPLICANT: American Home Products Inc.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING AND TREATING SCHIZOPHR  
; FILE REFERENCE: AM101228L  
; CURRENT APPLICATION NUMBER: US/60/427,194  
; CURRENT FILING DATE: 2002-11-19  
; NUMBER OF SEQ ID NOS: 234  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 154  
; LENGTH: 752  
; TYPE: prt  
; ORGANISM: Homo sapiens  
US-60-427-194-154

Query Match  
Best Local Similarity 5.4%; Score 126.5; DB 7; Length 752;  
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;



Db 530 QGDHSLVQKKLNQENRNDLG----REIENLTKELEERYRHS-----PSL 575  
Qy 376 LREMSSEONGDALSTIFNAGLNHLDOVLLPEVSTPTVGIESPOTQAPKPSVSDPSVEE 435  
| : : : : :  
Db 576 -----NGRISD-----PQVFSKEVQTEAVNNEPPDYKSLIP-----LER 610  
Qy 436 HVSQTSPTVQSKODKPOSSATSALS 463  
| : : : : :  
Db 611 AVINGQLYESEENODEDPNDEC-SVLSF 637

## RESULT 8

US-09-724-676a-58406  
; Sequence 58406, Application US/09724676a  
; GENERAL INFORMATION:  
; APPLICANT: Comugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Comugen  
; CURRENT APPLICATION NUMBER: US/09/724,676a  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 58406  
; LENGTH: 796  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676a-58406

Query Match 5.4%; Score 126.5; DB 5; Length 796;  
Best Local Similarity 21.9%; Pred. No. 0.018;  
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

Qy 5 KHIQIOLSKPLS-----DSSI-----CGVYLKLEKSAFRLPNEFNVAOTALK 48  
| : : : : :  
Db 204 KLEVERLSKRIMALLEKLEDAFNKSKOECYSLCN--LEKERMTTKOLSOELESKVRIKE 261  
Qy 49 LSONPSADERD--ALQEAFLNKWKILS-----DSLVOFSKTTTROIELISWFAOFL 99  
| : : : : :  
Db 262 LEAIESRLKTEFTFKED-LTKLKLTVMFVDERKTMSEKLTEDKQLQAASSQLQVEON 320  
Qy 100 LQTT-----LESANSLLEMLADLSEKHMHLNVLPEVETLKSDDDKGKEREQADAKVKA 154  
| : : : : :  
Db 321 KVTTVTEKLEIEETKRLKSKTDVEEKM-----SVTKERDDLKNKLKA- 363  
Qy 155 POLVGDSESSILYAPVLOLPLVEGVTFFDFOAER--KGEISOLKSMULTTVAQERA 211  
| : : : : :  
Db 364 -----PEEGNDLSRVNML-----KNRLOSLEAIEKDFLKNKLQDSKSTTALHQNKK 414  
Qy 212 IOFKMENARCVYQDLRLSAL-----VSTKCHSLGOSSTN-FGFAKSLITRVENALVHL 264  
| : : : : :  
Db 415 IKELSGEVERLKLKDKMAIEDDLMKTEDEYETTLERRRANERDKAQLSKELHVKMEL 474  
Qy 265 SGIKLAPKAA-----KTVEOEVAESSVSEGELPSHMDTKHIERIP--MASEQAQTVS 315  
| : : : : :  
Db 475 AKYKLAETKETSHEQWLFKRLQEEBAKS-----GHLSREVDALK-EKIHETMATEDLICH 529  
Qy 316 OHLHAGNLSELGNLNNMNDLAFHLLREVSDFROSEPHSPISFLEKAIIRMGYLSLPET 375  
| : : : : :  
Db 530 QGDHSLVQKKLNQENRNDLG----REIENLTKELEERYRHS-----PSL 575  
Qy 376 LREMSSEONGDALSTIFNAGLNHLDOVLLPEVSTPTVGIESPOTQAPKPSVSDPSVEE 435  
| : : : : :  
Db 576 -----NGRISD-----PQVFSKEVQTEAVNNEPPDYKSLIP-----LER 610  
Qy 436 HVSQTSPTVQSKODKPOSSATSALS 463  
| : : : : :  
Db 611 AVINGQLYESEENODEDPNDEC-SVLSF 637

RESULT 9  
US-09-724-676-58407  
; Sequence 58407, Application US/09724676  
; GENERAL INFORMATION:

; APPLICANT: Comugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Comugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 58407  
; LENGTH: 810  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-58407

Query Match 5.4%; Score 126.5; DB 5; Length 810;  
Best Local Similarity 21.9%; Pred. No. 0.019;  
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

Qy 5 KHIQIOLSKPLS-----DSSI-----CGVYLKLEKSAFRLPNEFNVAOTALK 48  
| : : : : :  
Db 218 KLEVERLSKRIMALLEKLEDAFNKSKOECYSLCN--LEKERMTTKOLSOELESKVRIKE 275  
Qy 49 LSONPSADERD--ALQEAFLNKWKILS-----DSLVOFSKTTTROIELISWFAOFL 99  
| : : : : :  
Db 276 LEAIESRLKTEFTFKED-LTKLKLTVMFVDERKTMSEKLTEDKQLQAASSQLQVEON 334  
Qy 100 LQTT-----LESANSLLEMLADLSEKHMHLNVLPEVETLKSDDDKGKEREQADAKVKA 154  
| : : : : :  
Db 335 KVTTVTEKLEIEETKRLKSKTDVEEKM-----SVTKERDDLKNKLKA- 377  
Qy 155 POLVGDSESSILYAPVLOLPLVEGVTFFDFOAER--KGEISOLKSMULTTVAQERA 211  
| : : : : :  
Db 378 -----PEEGNDLSRVNML-----KNRLOSLEAIEKDFLKNKLQDSKSTTALHQNKK 428  
Qy 212 IOFKMENARCVYQDLRLSAL-----VSTKCHSLGOSSTN-FGFAKSLITRVENALVHL 264  
| : : : : :  
Db 429 IKELSGEVERLKLKDKMAIEDDLMKTEDEYETTLERRRANERDKAQLSKELHVKMEL 488  
Qy 265 SGIKLAPKAA-----KTVEOEVAESSVSEGELPSHMDTKHIERIP--MASEQAQTVS 315  
| : : : : :  
Db 489 AKYKLAETKETSHEQWLFKRLQEEBAKS-----GHLSREVDALK-EKIHETMATEDLICH 543  
Qy 316 OHLHAGNLSELGNLNNMNDLAFHLLREVSDFROSEPHSPISFLEKAIIRMGYLSLPET 375  
| : : : : :  
Db 544 QGDHSLVQKKLNQENRNDLG----REIENLTKELEERYRHS-----PSL 589  
Qy 376 LREMSSEONGDALSTIFNAGLNHLDOVLLPEVSTPTVGIESPOTQAPKPSVSDPSVEE 435  
| : : : : :  
Db 590 -----NGRISD-----PQVFSKEVQTEAVNNEPPDYKSLIP-----LER 624  
Qy 436 HVSQTSPTVQSKODKPOSSATSALS 463  
| : : : : :  
Db 625 AVINGQLYESEENODEDPNDEC-SVLSF 651

RESULT 10  
US-09-724-676-58408  
; Sequence 58408, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Comugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Comugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 58408  
; LENGTH: 810  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-58408

Query Match 5.4%; Score 126.5; DB 5; Length 810;  
Best Local Similarity 21.9%; Pred. No. 0.019;





US-10-309-851-12

|                           |        |                  |             |              |
|---------------------------|--------|------------------|-------------|--------------|
| Query Match               | 5.48;  | Score 126.5;     | DB 6;       | Length 1135; |
| Best Local Similarity     | 21.98; | Pred. No. 0.032; |             |              |
| Matches 111; Conservative | 77;    | Mismatches 197;  | Indels 123; | Gaps 24;     |

[illegible]

```
Search completed: March  2, 2003, 05:41:33
Job time : 59 secs
```

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 03:23:43 ; Search time 39 seconds  
(without alignments)  
1141.288 Million cell updates/sec

Title: US-09-915-706a-2

Perfect score: 2343

Sequence: 1 MFLSKHQIEQLSKPLSDSI.....DRGSKQDOKQSSATSALSM 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID     | Description        |
|------------|-------|-------------|--------|--------|--------------------|
| 1          | 241.5 | 10.3        | 351    | AC3090 | conserved hypotet  |
| 2          | 241.5 | 10.3        | 351    | G98196 | hypothetical prote |
| 3          | 209   | 8.9         | 366    | D83350 | hypothetical prote |
| 4          | 132   | 6.5         | 2245   | T18278 | myosin heavy chain |
| 5          | 143   | 6.1         | 884    | T20405 | hypothetical prote |
| 6          | 141.5 | 6.0         | 344    | G83634 | hypothetical prote |
| 7          | 141.5 | 6.0         | 1510   | T16927 | hypothetical prote |
| 8          | 134.5 | 5.7         | 1179   | E98012 | hypothetical prote |
| 9          | 134   | 5.7         | 1051   | T18502 | hypothetical prote |
| 10         | 129.5 | 5.5         | 1133   | T22576 | apsb protein - Eme |
| 11         | 127.5 | 5.4         | 2471   | T42977 | large tegument pro |
| 12         | 126.5 | 5.4         | 1023   | AE1643 | ATP-dependent dsdn |
| 13         | 126   | 5.4         | 506    | S50914 | cell division cont |
| 14         | 125.5 | 5.4         | 1287   | T22235 | hypothetical prote |
| 15         | 125.5 | 5.4         | 1992   | A47297 | myosin heavy chain |
| 16         | 125   | 5.3         | 1194   | T37503 | probable chromosom |
| 17         | 124.5 | 5.3         | 550    | T40370 | hypothetical prote |
| 18         | 124.5 | 5.3         | 1940   | S04090 | myosin heavy chain |
| 19         | 124.5 | 5.3         | 1940   | A24922 | myosin heavy chain |
| 20         | 124.5 | 5.3         | 1964   | A59282 | nonmuscle myosin I |
| 21         | 123.5 | 5.3         | 867    | C71638 | DNA polymerase I ( |
| 22         | 123.5 | 5.3         | 1051   | T18351 | Imp1 protein - Myc |
| 23         | 123   | 5.2         | 1935   | A37102 | myosin beta heavy  |
| 24         | 123   | 5.2         | 1935   | S06006 | myosin beta heavy  |
| 25         | 123   | 5.2         | 2022   | T43214 | ovtl protein - nem |
| 26         | 122.5 | 5.2         | 2469   | H36812 | hypothetical prote |
| 27         | 122   | 5.2         | 1934   | I48153 | myosin heavy chain |
| 28         | 122   | 5.2         | 1934   | E89066 | protein H05009.1   |
| 29         | 122   | 5.2         | 2109   | T33247 | hypothetical prote |

|    |       |     |      |   |        |                    |
|----|-------|-----|------|---|--------|--------------------|
| 30 | 121.5 | 5.2 | 1179 | 2 | G95144 | conserved hypotet  |
| 31 | 121.5 | 5.2 | 3225 | 2 | I52300 | giantin - human    |
| 32 | 121.5 | 5.2 | 3259 | 1 | A56539 | giantin - human    |
| 33 | 120.5 | 5.1 | 999  | 2 | JC5278 | oxygen regulated p |
| 34 | 120.5 | 5.1 | 1938 | 2 | A59293 | skeletal myosin he |
| 35 | 120   | 5.1 | 650  | 2 | S44806 | F1059.6 protein -  |
| 36 | 120   | 5.1 | 828  | 2 | F96535 | hypothetical prote |
| 37 | 120   | 5.1 | 1034 | 2 | T32297 | hypothetical prote |
| 38 | 119.5 | 5.1 | 1066 | 1 | A48669 | kinasin-related pr |
| 39 | 119.5 | 5.1 | 1690 | 2 | T13030 | microtubule bindin |
| 40 | 119   | 5.1 | 1023 | 2 | AE1280 | ATP-dependent dsdn |
| 41 | 119   | 5.1 | 1356 | 2 | S32763 | kinectin 1 - human |
| 42 | 119   | 5.1 | 1935 | 2 | A59286 | myosin heavy chain |
| 43 | 119   | 5.1 | 2104 | 2 | T38774 | myosin-3 heavy cha |
| 44 | 118.5 | 5.1 | 1920 | 2 | A53188 | pericentrin - mous |
| 45 | 118.5 | 5.1 | 1937 | 2 | I38055 | myosin heavy chain |

#### ALIGNMENTS

##### RESULT 1

AC3090 conserved hypothetical protein Atu4343 [imported] - Agrobacterium tumefaciens (strain

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #extl\_change 11-Jan-2002

C:Accession: AC3090

R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woc

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuyavlin, T.; Levy, R.; Li, M.; MCCI

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kan

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AC3090

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-351 <KUR>

A:Cross-references: GB:AE008689; PIDN:ANL45137.1; PID:g17742810; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu4343

A:Map position: linear chromosome

Query Match 10.3%; Score 241.5; DB 2; Length 351;

Best Local Similarity 23.6%; Pred. No. 2.3e-08;

Matches 96; Conservative 64; Mismatches 162; Indels 85; Gaps 15;

|    |     |   |     |
|----|-----|---|-----|
| QY | 5   | KHOIEQLSKPLSDSDICGVYTLKLE---KSAFRPLRNFENVAQRTALRKLSPNSADEBDAL | 61  |
| DB | 7   | KRDIEFLGN-----CGDNIRNDSRTREIYRIKDRNRAR-----AEERAA5            | 48  |
| QY | 62  | QEAQL---NKKILSDSLVEQFSKTRTDIELISGFVAAGFLDTTLESANSLLEMLADLS    | 118 |
| DB | 49  | PODLKLSSWDSVSNLGLDIYSESKDVELLMLAASRL- RGFHGLREIYELCGDLF       | 107 |
| QY | 119 | EKKHDLNPLVLPVETTLKSDDDKGERQADAKKAFQVLGVDSSESSILYAPVQLPLV-     | 177 |
| DB | 108 | YNHMDSLRSI-----SDND-----BEKRAFPGGLNGISSEGT-LVQPLRLSLIP        | 152 |
| QY | 178 | -----GEVTFDFQSAERKGEISOLKSLTTTVAQERFAIOFKMKNARCTQIDRLSALV     | 233 |
| DB | 153 | GKGFGEHSLMDFOLAQPNRSKRREELRYLASEGVAAAMSHLAAVTCLSSDPAITAVL     | 212 |
| QY | 234 | STACHSLGSGSTNFGFAKSLTLRVENALVHLSGIKLAPKAERATV---EQEVAESSVSEG  | 290 |
| DB | 213 | SERCGQAAPPSSN-----IRNTLIEAA-----AAIRTLGGRDGEPAP-----          | 249 |
| QY | 291 | ELPSHMDTKIERIPMASEAOQIVYSHLHAGNLSELGNLNNMNDLAFLHLLREVSDFRQ    | 350 |
| DB | 250 | -----VEQTPALA--AGTDESGQSAARTSPASPEGISSRDEAFETLLSVARVFR        | 297 |



|    |                       |  |                    |       |                 |                     |
|----|-----------------------|--|--------------------|-------|-----------------|---------------------|
|    | Query Match           | 6.5%   | Score 152;         | DB 2; | Length 2245;    |                     |
|    | BEST Local Similarity | 23.6%;   | Pred. No. 0.2;     |       |                 |                     |
| Db | Matches               | 81;  | Conservative       | 57;   | Mismatches 143; | Indels 62; Gaps 13. |
| QY | 23                    | VYLKLESAFPLRNEFNVAOTALRKLSQNSADERDALQEQCLNKKWIKLSDSYEOLF82   | :                  | :     | :               | :                   |
|    |                       | : : : : : : : : : : : : : : : :                              | :                  | :     | :               | :                   |
| Db | 964                   | IQLREANSLTVOEQNNKLQEKLEELQMRLLSEAKRRQQ---                    | LDDQKYKSPTTTSSELLS | 1020  |                 |                     |
| QY | 83                    | TTRDIEL-TSNFVAAQFLIDTTLESANSL-EWLADLSEK----                  | HMDHLNVLPEVTLK     | 135   |                 |                     |
|    |                       | : : : : : : : : : : : : : : : :                              | :                  | :     | :               | :                   |
| Db | 1021                  | NNDHLELOLSEIQLYOGLDNSNOSSQJLSECLSELBQTQDLHSSKLKLEMLSD        | 1080               |       |                 |                     |
| QY | 136                   | SDDDKGREQADAKVKAFQVLVDGSEESSILYAVALPLVGEXTF----              | PDPOSAER-          | 190   |                 |                     |
|    |                       | : : : : : : : : : : : : : : : :                              | :                  | :     | :               | :                   |
| Db | 1081                  | QHDSTIERKQSFNETEQOLOOPKOOSESELRSKLSTTQ----                   | QUDFNKQEFDRLSQEND  | 1135  |                 |                     |
| QY | 191                   | -----KGELSQKSMLTTPVAQERFALOFGMENAKRCVTOIDLRLSALTSTCHSIGSOS   | 244                |       |                 |                     |
|    |                       | : : : : : : : : : : : : : : : :                              | :                  | :     | :               | :                   |
| Db | 1136                  | TDNTNNGLEIQOKK-ANSTLEEEDFSLSGINDINEROVLELRDENOLIKELDSLGGQS   | 1194               |       |                 |                     |
| QY | 245                   | TNF--GFA-----KSLLTRVENALVHLGILAPKAEEK-----                   | IVE                | 279   |                 |                     |
|    |                       | : : : : : : : : : : : : : : : :                              | :                  | :     | :               | :                   |
| Db | 1195                  | SOFQGAALLEKQOELOLVQEOESEQJIKLSSEKSGEEEAKKQINOLETELDTHKSKXIOQ | 1254               |       |                 |                     |
| QY | 280                   | QEVAESSVS-----EGELPSHMDTK-----HIERIPMASEQQ                   | 312                |       |                 |                     |
|    |                       | : : : : : : : : : : : : : : : :                              | :                  | :     | :               | :                   |
| Db | 1235                  | LQTLBDSNEKITKLGKLEEVODEKKQLOOLEIERIKOSQSYE                   | 1297               |       |                 |                     |

| Query Match | Best Local Similarity                                     | Matches | Score | DB 2: | Length |
|-------------|---|---------|-------|-------|--------|
| 27          | LEKSAFRLREHFNVAOALRKLSQPSADERDALOEACLNKMW                 | 6.18;   | 143;  | DB 2: | 884;   |
| 218         | LRPDAVYATARTTRRLALSTALSLR                                 | 20.7%;  | 99;   | 165;  | 130;   |
| 83          | TTROIELISVVAOFLDITLLESANSLLEWLADLSEKHMCHLNPVLYETLKSDDDKGK | 142     |       |       |        |
| 269         | LTRPEVDLVETTSKLAVEAPETTTQEEFTTDBASE                       | 317     |       |       |        |
| 143         | E-RQGAADKAKAFQOLVQDSESSILTVAPVLO                          | 186     |       |       |        |

[illegible]

RESULT 6

G83634

hypothetical protein PA0082 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C:Species: *Pseudomonas aeruginosa*

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: G83634

R:Stover, C.K.; Plam, X.O.; Erwin, A.L.; Micooguchi, S.D.; Warrenner, P.; Hickey, M.J.; L

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L

Loiry, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: G83634

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-344 <STO>

A:Cross-references: GB:AE004447; GB:AE004091; MID:g9945902; PIDN:AAG03472.1; GSPDB:GN

C:Genetics:

A:Experimental source: strain PA01

A:Gene: PA0082

Query Match 6.08; Score 141.5; DB 2; Length 344;

Best Local Similarity 21.18; Pred. No. 0.067;

Matches 89; Conservative 70; Mismatches 144; Indels 119; Gaps 18;

QY 15 LSDSDIGCVYLKLEKSNRPLRNFNVAQTARLTSQ-NPSADERDADLQEACLKKWTLS 73

DB 11 VSPDSGCGDDL-----ETDAFLELEIRIAQQPERONGDAVLAPEPPMPRRV 58

QY 74 DSLYQFSKTRDIELLSWFAAOFLLDTTLESANSLSEWLADESEKHMDLNLVPEV 133

DB 59 ALASLELFR-SKDLRVANLLLSQNSVALD-GLDGLADGLLIVRELIGQYWDGYVL----- 111

QY 134 LKSDDDKGEREQADAKYKAFQVLVGDSSESSILYAPVLQ-----PLVGEVTF--PDFQS 187

DB 112 LDADD-----NDPTFRINALGLVAE-----PLIQVMAIPIVYRSRFGPVNIRA 157

QY 188 AERGEISQLKSMLTTPYAOERFALQFMENAKRCYVQDLDSLALVSTKCKSGLSQSTNF 247

DB 158 ALNAAQLQRFAS-----ETLSPEQIAGAFPADADA-----DALATRRALDQAQEHAL 204

QY 248 GFAKSLTRRVNA-----LVHL-----SGIKLAPKAEAKTVQOEVAES 285

DB 205 AIESGVAVMERVGSAGQGLDGLRLQALLRQALQYFDLYPGQAGESIALPAGEA-VADQYGA 263

QY 286 SVSEGEELPSHMDTKHIERIPMASDAQOTVSOHLHAGNLSELGNLNNNRDLAFHLREVS 345

Db 264 PVAAPAA-----PRAS-----CEIA-----NREDVLRQLRL 293

Qy 346 DYERQSEPHSPISFLLEKAIKRWGYSLEPELLREMSQONDALSTINAGLNLDVLL 405

Db 294 EYVVRHEPSSVPVLLKRAKTLVTADPAEIVRNLPD-----GISQFETLRG 340

Qy 406 PE 407

Db 341 PE 342

RESULT 7

116927

hypothetical protein T23F2.2 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T16927

R:Du, Z.

submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid T23F2.

A:Reference number: Z18608

A:Accession: T16927

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1510 <DU2>

A:Cross-references: EMBL:U03649; NID:g1049370; PID:g1049372; PIDN:AAA80384.1; CESP:T23F2

C:Genetics:

A:Gene: CESP:T23F2.2

A:Introns: 20/2; 81/1; 119/2; 211/3; 236/3; 290/2; 344/1; 396/3; 439/3; 465/3; 556/3; 60

Query Match

Best local similarity 20.0%; Score 141.5; DB 2; Length 1510;

Matches 112; Conservative 89; Mismatches 168; Indels 191; Gaps 26;

Qy 9 EQLSKPLSDSICGVYIKLEKSAFRLRNFNNAOTAKRLSONPSADDERDLQEA 68

Db 619 QOLSEHSD-----FLSTQOT-LHDLKKEYSAKNTTL-----VDKREVEEILLAK 663

Qy 69 WKILSDSLVEQFSK-----TRDIELISMFAVAOFLDPTLESANSLLEMLADISEKHMDH 124

Db 664 TELV-DALVTQOLENIRKQDREL-----LKQSERDQYKKSLEEMFTIAEK----- 708

Qy 125 LNPVLPVETLKSDDDKGEREQADAKVK--AFQO-----LVGDS----- 161

Db 709 -VPLLEAEIILQLSKDK--NEITARLKHQEYFEDELAKLNDSMIKKERDDYLTENI 763

Qy 162 --EESSLVAPVQLPLVGEVTFPDSAR-KGEISQLKSMLTYYAQERFALQFMEN 218

Db 764 RANESML-----ERKLEISGLKDLLENOKMOAHLQ--KSEL 798

Qy 219 AKRCVTOLDRLSALVS-----TKCHSLSGSQSTNFGFAK----- 251

Db 799 EKRLLSDIHVSQOSRVNHSQRDVECOAIPROINKYVGCKPKNKNETIIEKALDEN 858

Qy 252 -----SLITRVENALVHL-----SGIKLAPKADEAKTVEQVAESS 286

Db 859 EERLRICKAELETTTRRVTVLQQLVSIIOQSSQKIKKRIAYVEDNSKNVTHTEDLESK 918

Qy 287 VSEGEPLSHMDTKHIERIPMASEQOFTVSOHLHAGNLSLGN-----LNNMNR 334

Db 919 MKEVEL--KNTLEMERID-SLEAEFVAASIEKSRIOQLVNEEDNLKOKIDNDMSVYSK 974

Qy 335 DLAF-----HLREVSDFROSEPHSPISFLLEKAIKRWGYSLEPE-LTREMSEON-CD 386

Db 975 EKQWLQWRISLLEKDNELQKLOPSS-----EKSLK-----SLNKGLTRKTSBEPFGD 1024

Qy 387 ALSTFINAGLNLDVLLPEVSTPTVIGIESPQ-----TPQAKPSVSDPR 431

Db 1025 DMSTEGGASSTESADPMSVTFAPVLSKSPQFOLADVLNLVRSDLQVLTIEIEPE 1084

Qy 432 SVEEHVQSPTSPVQNSKQDQ 451

Db 1085 AAKQEPQMSLEKSTKNVQ 1104

RESULT 8

E98012

hypothetical protein smc [imported] - *Streptococcus pneumoniae* (strain R6)

C:Species: *Streptococcus pneumoniae*

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C:Accession: E98012

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Buggett, S.; Dehoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; Mcahren, S. Y., P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.

A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: E98012

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1179 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99929.1; PID:g15458753; GSPDB:GN00174

C:Genetics:

A:Gene: smc

Query Match

Best local similarity 21.6%; Score 134.5; DB 2; Length 1179;

Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

Qy 5 KHQEQSLKPLSDSICGVYIKLEKSAFRLRNFNNAOTAKRLSONPSADDERDLQEA 64

Db 674 KPELEQLQKELADE--ASLGSEALKTLQDQMAALTELEAK--SGEQARIQEQ 727

Qy 65 CLNKWKILSDSLVEQFSKTRDIELISMFAVAOFLDPTLESANSLLEMLADISEKHMDH 124

Db 728 GLS-----LAVQQTQSQVBELETL-WKLOEE-EIDRLSEG-----DWQAD-KEQQES 772

Qy 125 L-----NPLVPEFLKSDDDKGEREQADAKVKAFFOLVGDSESSSLIYAPVQLPL 176

Db 773 LATIASERKQMLEAEIERIKSNKNAIQERYO-----NLQEEVAQARLLKTEL 818

Qy 177 VGEVTFPDSARKEKGEISQLKSMLTYYAQERF-AIQEMENAKRCVQOLRLSALVS 234

Db 819 QGQKR-YEVADIERLG-----KEIDNINIEQEEIQRMLQEKVDNLEKVDTE-----LLS 866

Qy 235 TKCHSLSGSQSTNFGFAKSL-LTRVENALVHLG-----IKLAPRAEAKTV 278

Db 867 QQAESKTQKTNLQOGLIRROFELDIEGQLDIALSHLDQARQONEWIRKQTRAEAK-- 924

Qy 279 EOEVASESSVSEGEPLSHMDTKHIERIPMASEQOFTVSOHLHAGNLS--ELGNLNNMRDL 336

Db 925 KEKVSERL-----RLHQQLTDQOISYTALEKAKHELENLNLAEQVQDLEKALRSL 977

Qy 337 AFHLREVSDFROSEPHSPISFLLEKAIKRWGYSLEPELLREMSQONDALSTINAGLNLDVLL 393

Db 978 GPVNLAEIDQY--EEVHNRLDPL--NSQRDILSANKLLLETTEKNDVEKRFKSTFE 1032

Qy 394 AA-----GLNHLQVLLPEVSTPTVIGIESPQEQAK 424

Db 1033 AIRESFVTFKQMFQGGQAD-LITTEGDLTLGAVEISVQPPCK 1074

RESULT 9

T18302

apsA protein - *Emeticella nidulans*

C:Species: *Emeticella nidulans*, *Aspergillus nidulans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000

C:Accession: T18302

R:Stelmann, R.; Slevens, N.; Galetzka, D.; Robertson, L.; Timberlake, W.E.; Fischer, M.O. Microbiol. 30, 831-842, 1998

A:Title: Increased nuclear traffic chaos in hyphae of *Aspergillus nidulans*: molecular

A:Reference number: Z18870; MUID:99140441; PMID:10094631

A:Accession: T18302

A:Status: preliminary; translated from GB/EMBL/DBJ



```

Db 758 -----AESHQWNEFPASPIIPQLOLDEIKTEVQRTAKOQOETLNOJLSVOS 808
QY 157 LVGSEESIILYAPVLOL-----PLVGE---VFFPDQSAREKGEISO--LKSMLTTT- 204
      |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 809 LLSQAEOSSTSTIPLOHYTQAGTLVGESEKNETFESJRDYOKLSTSEEFKLNINST 868
QY 205 -----V 205
      |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 869 LLNIQTOIKDISILLSNQWYQSETIKHAFLEKSNLTIGEAIOJINIKKTYDTLOPILI 928
QY 206 AOERPAIOFK-----MENAKRCVYQDLDRLSALYSTKCHS 239
      |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 929 AVKRRLSAKEREKSDNIYEIIYTTLSIGSLLSNPETIALKDYLKSIDTLPRKATVSKP 988
QY 240 LGSOSTNF--GFAKSLTRVE-----NALVHLSGIKLAPKREA 275
      |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 989 LKRELIYNNMKLOKDLKQLOKOLENMKMEVEYTFASTPGRDVNF-----IQSPSTKA 1043
QY 276 KTVDEQVAESSVSEGELPSHMDTKHIERIPMAEQAOQTVOGHLHAGNLSGCLNNMNRD 335
      |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 1044 RKYAEVLEKDOIOE-----MD-----IDVSP E--SIIDNIRKANGOKAMWKIQSAQD 1089
QY 336 LAFHLRLREVSQYFROSEBHS--PISFL-----LEKARWVYLSLEPL-----LRPMSEQ 383
      |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 1090 LNFSTL--PADMWLSIAKERYSDOSLSFTTIGPTLTKLKEEVFVSIOQNKDAKLSL--P 1146
QY 384 NGDALST-----IFNAAGLN--HLDQVLLPEVST--PTVGIESPOTPOAKPSVSPRSY- 433
      |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 1147 NGPAPTSKELMDIHIEYEDNVNPFHLKTIINLPKISTIAHTIGHELMSLOALNSKTLPEAVV 1206
QY 434 -----EHHVSOTS-----PYDTOSKODQ 451
      |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 1207 GTPLEOHAKFSCKMFKTLEATVMDHDOYDFRKIKLE 1241

```

RESULT 12  
AE1643  
ATP-dependent dsDNA exonuclease SbcC homolog sbcC [imported] - *Listeria innocua* (strain  
C)Species: *Listeria innocua*  
C|Date: 27-Nov-2001 #sequence.revision 27-Nov-2001 #text.change 27-Nov-2001  
C|Accession: AE1643  
R|Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Esthl, H  
D.; Jones, L.M.; Kaerst, U  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueder, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A>Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11079669  
A:Accession: AE1643  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1023 <G|A>  
A:Cross-references: GB:AL592022; PIDN:CA069617.1; PID:g16414173; GSPDB:GN00178  
A:Experimental source: strain Cl1p11262  
C:Genetics:  
C:Gene: sbcC

|                           |        |                 |             |              |
|---------------------------|--------|-----------------|-------------|--------------|
| Query Match               | 5.48;  | Score 126.5;    | DB 2;       | Length 1023; |
| Best Local Similarity     | 20.38; | Pred. No. 3;    |             |              |
| Matches 102; Conservative | 82;    | Mismatches 192; | Indels 127; | Gaps 21      |

```

0Y      8  IEOLSKPLSDSDISGCVLLKLEKSAFRLRNBNFVAQCALRKLKSPNSFADBRDALQECJCN 67
      |||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      284  IEEVAK-----LEAEKDFYOVYIANLEAKKRAMSIRSDODALCIRLAKQOLEEAVN 332
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
0Y      68  KMKILIS--DLYEQFSKRTTRDIELIMFVAAOPLDPTTLEASANSLEMLADISEK---- 120
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      333  TEKOVAHEVELITQFSNAKQKEAL---AEQ---EAVLEANKRTTLEQEEHEPKITEL 385
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
0Y      121  -----HW-----DHNLPLPYE-----TLKSDDDQKGEREQCADKVAAPQLQV 159
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      386  ETVYIQRRAETINKEADRLERKVMKTEQETIAELQSMETRLGEOINAEI---ALLFAIN 442
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

```

0Y 160 DSESSILYAVLDPLDVGEXTFPDFOSAEKKGISOLKMSLMTTVAQEFALO--FKME 21
Db 443 KRTTAREALIEK--NOELVNNKMKKMDLWNKOKOTEEOTLKOLLT-----EKLAETMIKOE 495
0Y 218 NAKRCVOTDLDRLSALV--STKCHSLGS-----QSTNPFPAKSLTTRVENALVHLGSIKLAP 271
Db 496 ESNLQOEBAATLALHLHGDCACRPVCSVSHQLAAYEGESASLTTLFE----- 542
0Y 272 KAEATVAEQEVAA----ESSVS--EGELPSHMDTKHIERIPMASEQOATVSOHLHAGNLSE 325
Db 543 -AKAKLHEKOLAIINEVEKSIISOLEMOLAEMADIDVDVLVAVERKLAA-- 588
0Y 326 LGNLNNMRDLAFHLIRVSDYFFQSEPHSPISLLEKALRWGYSLSPELLREMSBONG 385
Db 589 -----NROLAKNLTQIDINOLOTVNAOKENIGALAE-----TLKNKOKLELETENK 633
0Y 386 DAL-----STFNAGLNLHLDVLLPEVSTPIVGIESTPOTQAPKSVSDPSVEBHV 438
Db 634 TALOVESTHGEVOLLSSGKSLSTLEBOALPADLDKTT-----FDKKKNELSN--SIRTHLE 685
0Y 439 QTSPIVDTQSKODKPOSSATSAAL 461
Db 686 QAEQVDKIFREAEKETTRPLESTL 708

```

RESULT 13  
S50914  
cell division control protein CDC37 - Yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: Cdc37 start control protein; protein Y9489.03; protein YDR168W  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 10-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999  
C:Accession: S50914; B26372  
R:Olliver, K.; Harris, D.  
Submitted to the EMBL Data Library, January 1994

A:Cross-references: EMBL:Z47813; NID:9642274; PIDN:CMAB7799.1; PID:9642277; MIPS:YDR168w  
R:Refugson, J.; Ho, J.Y.; Peterson, T.A.; Reed, S.I.  
Nucleic Acids Res. 14, 6681-6697, 1986  
A>Title: Nucleotide sequence of the yeast cell division cycle start genes CDC28, CDC3  
A:Reference number: A93635; MUID:86313926; PMID:3018676  
A:Accession: B26372  
A:Molecule type: DNA  
A:Residues: 58-168, 'D', 170-506 <FER>  
A:Cross-references: EMBL:X04288; NID:93492; PIDN:CA427836.1; PID:93493  
A>Note: the authors translated the codon GAC for residue 112 as Ala  
C:Genetics:  
A:Gene: SGD:CDC37  
A:Cross-references: SGD:S0002575; MIPS:YDR168w  
A:Map position: 4R  
C:Keywords: cell cycle control

|                       |        |                  |                 |             |
|-----------------------|--------|------------------|-----------------|-------------|
| Query Match           | 5.4%;  | Score 126;       | DB 2;           | Length 506; |
| Best Local Similarity | 19.1%; | Pred. No. 1.2;   |                 |             |
| Matches               | 90;    | Conservative 71; | Mismatches 161; | Indels 148; |
|                       |        |                  | Gaps            | 17;         |

```

OY 15 LSDSICGVYLKLEKSARPLRNEENNQAOTLRLKLSQNPASDEADALQOEACLNKWKILSD 74
    ||||| :||| :|||
Db 13 LSDSDVEVHEHNVDKKSP-----IKW--OQ 36
OY 75 SLVEFSTTDDIELISFVAQAQFLDOTTLESANSLLEMLADLSEKHMDHNLPLVPETL 134
    ||||| :||| :|||
Db 37 SIHEQRFERNODIKLETLQVLMYSHLNKRVDRI-----LSNPESSITLDELPAATKFLMA 90
OY 135 KSDDDKGERQADAKVAFAFQVLGDSSESSILYAPVQLPLVGEVFFPDQSAERKEGEI 194
    ||||| :||| :|||
Db 91 NEDKMEKSGENVDPPEIATYNEMVEDLFE-----QLA-----KDLDKEGKD 131
OY 195 SOLKSMLETTVAQEFALQFEMENAKRCVYDLDRLSALVYKCHSLGSGQSTNPFASLL 254
    ||||| :||| :|||

```

Db 132 SKSPSLIRDAILKHKRAKIDSVTEAKK-----KIDELYKKNNAHISEDINTGPDSPFM 185  
Oy 255 TRVENALVHLSGI---KLAPKAERAKTVEOEVAESSVSE-----GEL 292  
Db 186 NKOGAKLEAPFPSPALSSAES-NILNKLAKSVPPQFFIDPKDDPKMLAKETEEFGKI 244  
Oy 293 PSHMDTKH---IERIPMASEQ-----AQVSOHLHAG----- 321  
Db 245 SINEYSKQKFLLEHLPIISEQOKDALMMKAFEYOLHGDDKMTLOVHOSELMAYIKEIY 304  
Oy 322 -----NLSELGNLNNNNRDLAFHLLEVSDFPROSEPHSIFSLKAIWGYLSLP 373  
Db 305 DMKKIPLYLPMELSNVINM-----FFEKV--IFNKDKPMGSEFL--RSVQEKFLHIQ 353  
Oy 374 E---LLREMSSEONGDALSTIFNAGLNHLDOVLLPEVSTPYVGLSESPQ 419  
Db 354 KRSKILQOEEMDESNAEGVETI---QLKSLDSTLEVLNLPDPSKDP 399

## RESULT 14

T22235  
hypothetical protein F45G2.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22235  
R.Lindsay, S.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19535  
A:Accession: T22235  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1287 <MIL>  
A:Cross-references: EMBL:Z93382; PIDN: CAB07612.1; GSPDB: GN00021; CESP: F45G2.3  
A:Experimental source: clone F45G2  
C:Genetics:  
A:Gene: CESP:F45G2.3  
A:Map position: 3  
A:introns: 59/2; 117/3; 153/3; 180/2; 233/2; 341/3; 393/1; 424/2; 551/3; 664/3; 734/3; 8

## Query Match

Best Local Similarity 5.4%; Score 125.5; DB 2; Length 1287;  
Matches 90; Conservative 79; Mismatches 167; Indels 85; Gaps 18;

Oy 7 QTEQLSKPLSDSICGVYKLEKSAPFRPNEFNVAQTALRKLSQNPASADERALDEACL 66  
Db 259 EIEELNRNLKAESSENYENAKKASELERLR-----EKEKWEDEKERRQMAENDA-- 310  
Oy 67 NKMKILSDLYEOFSTTRDIELISFVAQFLDITLESANSLLEMLADLSEKHWDLN 126  
Db 311 NQHKOTERKLQDQISTLQTDFEKVSQORKAQOEOMNAELVDEVASFQKAKERADEQKKTL- 369  
Oy 127 PVLVPVITLSDDDK-GKEEQADAKYKAFQVLVGDESESSILYAPVQLPRL-VGEVTFPD 184  
Db 370 ---VDLDSLDKDLAKERANNEHVKHKKLEGO-----LKATOTQLTLAKKHEED 419  
Oy 185 FQSAERKGEISQLSMLTTTVAQERFAIOFKME-NAKRCVTOURLSALSTCHSLSGQ 243  
Db 420 VQCKRRESIGELK-----LKAQGDANLISKLOAMLKICISRIEELDEDLLE- 467  
Oy 244 STNFGAKLLTRVENALVHLSGIKLAPKAERAKTVEOEVAESSVSEGLPESHMDTKHIER 303  
Db 468 -----RLRMKAERQFNEL-----RSEYEVLOEQMAEAS--GQLTAE--AHINK 507  
Oy 304 IPMASQOQTVSOHLHAGNLSE-----LGNLNNNR---DLAFHLLEAVSD 346  
Db 508 V--RAEVSNLRRDLQKRLNLEHAYISDLGNMOYATVNNLRNLSQOFSEFNCFEFLFL 565  
Oy 347 YFROSEPHSPIS-FLLEKAIWGYLSLPLELRMSSEONGDALST-----IFNAGL 397  
Db 566 FEFESFYDKRPINHLFOIRKTIK---SPVTYVIFHLLISRESVPTVASLCITAYQTIIFNGASL 622  
Oy 398 N 398

Db 623 N 623

## RESULT 15

A47297

myosin heavy chain form B, nonmuscle - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Feb-2001

C:Accession: A47297; A55441

R.Bhatia-Dey, N.; Adelstein, R.S.; David, I.B.

Proc. Natl. Acad. Sci. U.S.A. 90, 2856-2859, 1993

A:Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmu

A:Reference number: A47297; MUID:93219383; PMID:846400

A:Accession: A47297

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1992 <BHA>

A:Cross-references: GB:I09740; NID:9214623; PIDN:AAA49915.1; PID:9214624

A:Experimental source: XTC cells

A:Note: sequence extracted from NCBI backbone (NCBI:P128722)

R.Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.

J. Biol. Chem. 270, 1395-1401, 1995

A:Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p

A:Reference number: A55441; MUID:95138137; PMID:7836406

A:Accession: A55441

A:Status: preliminary

A:Molecule type: protein

A:Residues: 198-232 <KEL>

C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: nucleotide binding; P-loop

F:88-787/Domain: myosin motor domain homology <MMOT>

F:178-185/Region: nucleotide-binding motif A (P-loop)

## Query Match

Best Local Similarity 5.4%; Score 125.5; DB 2; Length 1992;  
Matches 107; Conservative 78; Mismatches 189; Indels 131; Gaps 20;

Oy 1 MPKSKQIOLSKPLSDSICGVYKLEKSAPFRPNEFNVAQTALRKLSQNPASADERA 60  
Db 929 LAIKQOEMLINDLE-----TRMEEEERNVLYLNKKKMTTHVQDLLE--QLDEEER 980  
Oy 61 LDQACLNK-----WKILSDLY--EOPSTTRDIELISFVAQFLDITLESANSL 111  
Db 981 AQLQLEKVTAEAKIKKMEEDILVLEDQNSKFLKEKKLLLEERIAESTQSALAEERAKNL 1040  
Oy 112 EWLADLSEKHWDLNVLVPVETLKSDDDKKEREQADAKYKAFQVLVGDESESSILYAPV 171  
Db 1041 AKLKNQOEKMSIDLE-----ERLKEEKTROELEKAKRK----- 1074  
Oy 172 LDPLVGEVTFPFGQSAERKGEISQLSMLTTTVAQERFAIOFKMEAKRCVTO----- 225  
Db 1075 ---LDGETTDFODQIAELQAOIEELQLL---AKKEEQLALARGDEEVLQKNNITLK 1126  
Oy 226 -LDRLSALVSTCHSLSGOSTNFGFA-----KSLTRVENALVHLSGIKLAPKA 273  
Db 1127 LVRELQAOIAELQEDLESEKASRNKAKKQKRDSELELAKTLEDTL-----DTAAQ 1181  
Oy 274 EAKTV-EQVEAESSV-EGELPSH-----MDTKHIERIPMASQOQTVSOHLHAGN 322  
Db 1182 ELRTKREOEVAELRKSIETETRNHEAOIEMRQORATALEEISEQEOAKRFKVNLEKN 1241  
Oy 323 LSELGNLNNNRDLAFHL-----LREVSDFRO-----SEPHSIFSLKAIWGYLS 371  
Db 1242 QS-----LESQNKELATEVSSLOQMKSEYKKKKLLEGQVOELHAKV---LE----- 1285  
Oy 372 LPELLRMSSEONGDALSTIFNAGLNHLDOVLLPEVSTPYVGLSESPQTPQAKPSVSDPR 431  
Db 1286 -GRLRADVYKSSKQLONELENVSS-----LLEAEKKKGIVL-----AKDVA 1326  
Oy 432 SVEEHVSQTSPPVDTQSKQDQKPOSS 456  
Db 1327 SMESQLODQOELLQETROKLANOSS 1351

Search completed: March 2, 2003, 05:06:12  
Job time : 52 secs

---

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 01:09:08 ; Search time 22 Seconds

(without alignments)  
872.888 Million cell updates/sec

Title: US-09-915-706A-2

Perfect score: 2343

Sequence: 1 MPLSKHQIEQLSKPLSDSI.....DTQSKQDQKPOSSATSALSW 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

# SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result No. | Score | Query Match | Length | DB ID        | Description        |
|------------|-------|-------------|--------|--------------|--------------------|
| 1          | 152   | 6.5         | 2245   | 1 MYSJ_DICDI | P54697 dictyostell |
| 2          | 126   | 5.4         | 506    | 1 CC37_YEAST | P06101 saccharomyc |
| 3          | 125   | 5.3         | 793    | 1 REGA_DICDI | Q23917 dictyostell |
| 4          | 124.5 | 5.3         | 1940   | 1 MYH3_HUMAN | P11055 homo sapien |
| 5          | 124.5 | 5.3         | 1940   | 1 MYH3_RAT   | P12847 rattus norv |
| 6          | 124   | 5.3         | 866    | 1 MYSP_SCHMA | P06198 schistosoma |
| 7          | 123.5 | 5.3         | 867    | 1 DP01_RICPR | O05949 rickettsia  |
| 8          | 123   | 5.2         | 1935   | 1 MYH7_HUMAN | P12883 homo sapien |
| 9          | 123   | 5.2         | 1935   | 1 MYH7_RAT   | P02564 rattus norv |
| 10         | 123   | 5.2         | 2022   | 1 ANTI_ONCVO | P21249 onchocerca  |
| 11         | 122.5 | 5.2         | 2469   | 1 TEGU_HSVSA | O01056 herpesvirus |
| 12         | 122   | 5.2         | 1934   | 1 MYH7_MESAU | P13340 mesocricetu |
| 13         | 120.5 | 5.1         | 999    | 1 OXRP_HUMAN | Q09411 homo sapien |
| 14         | 120.5 | 5.1         | 1938   | 1 MYH4_RABIT | Q28641 oryctolagus |
| 15         | 120   | 5.1         | 650    | 1 MI10_CAEEL | P34400 caenorhabd  |
| 16         | 119.5 | 5.1         | 1066   | 1 K161_DROME | P46863 drosophila  |
| 17         | 119   | 5.1         | 1935   | 1 MYH7_PIG   | P79293 sus scrofa  |
| 18         | 119   | 5.1         | 2104   | 1 MYH3_SCHPO | O14157 schizosacch |
| 19         | 118.5 | 5.1         | 1920   | 1 PCNT_MOUSE | P48125 mus musculu |
| 20         | 118.5 | 5.1         | 1937   | 1 MYH8_HUMAN | P13335 homo sapien |
| 21         | 117.5 | 5.0         | 539    | 1 YAH2_YEAST | P39705 saccharomyc |
| 22         | 117.5 | 5.0         | 1427   | 1 REST_HUMAN | P30822 homo sapien |
| 23         | 117.5 | 5.0         | 2116   | 1 MYS2_DICDI | P08799 dictyostell |
| 24         | 117   | 5.0         | 1679   | 1 Y109_YEAST | P40457 saccharomyc |
| 25         | 116.5 | 5.0         | 4385   | 1 YP73_CAEEL | O09222 caenorhabd  |
| 26         | 116   | 5.0         | 1616   | 1 P200_MYCGE | Q09429 mycoplasma  |
| 27         | 116   | 5.0         | 1966   | 1 MYSB_CAEEL | P02566 caenorhabd  |
| 28         | 115.5 | 4.9         | 2468   | 1 MABP_HUMAN | P46821 homo sapien |
| 29         | 115.5 | 4.9         | 3660   | 1 DMD_CHICK  | P11533 gallus gall |
| 30         | 115   | 4.9         | 1312   | 1 RA50_YEAST | P12753 saccharomyc |
| 31         | 114.5 | 4.9         | 955    | 1 K1NL_LEICH | P46865 leishmania  |
| 32         | 114.5 | 4.9         | 2230   | 1 G064_HUMAN | Q13339 homo sapien |
| 33         | 114.5 | 4.9         | 5327   | 1 ACE7_MOUSE | O9q420 mus musculu |

|    |       |     |      |              |                    |
|----|-------|-----|------|--------------|--------------------|
| 34 | 114   | 4.9 | 866  | 1 MYSP_SCHMA | O05870 schistosoma |
| 35 | 114   | 4.9 | 1940 | 1 MYH3_CHICK | P02565 gallus gall |
| 36 | 114   | 4.9 | 3672 | 1 LML2_CAEEL | Q21313 caenorhabd  |
| 37 | 113.5 | 4.8 | 671  | 1 CHEA_BACSU | P29072 bacillus su |
| 38 | 113   | 4.8 | 524  | 1 K2C4_MOUSE | P07744 mus musculu |
| 39 | 112   | 4.8 | 707  | 1 YJ9C_YEAST | P47166 saccharomyc |
| 40 | 112   | 4.8 | 804  | 1 MEA6_HUMAN | O15320 homo sapien |
| 41 | 112   | 4.8 | 1939 | 1 RECN_HUMAN | P13533 homo sapien |
| 42 | 111.5 | 4.8 | 522  | 1 RECN_HUMAN | O92180 helicobacte |
| 43 | 111.5 | 4.8 | 733  | 1 YH34_HUMAN | O9upw6 homo sapien |
| 44 | 111.5 | 4.8 | 1939 | 1 MYH1_HUMAN | P12882 homo sapien |
| 45 | 110.5 | 4.7 | 1252 | 1 RPOB_CHLPP | O929a0 chlamydia p |

## ALIGNMENTS

```

RESULT 1
ID      MYSJ_DICDI
AC      P54697;
DT      01-OCT-1996 (rel. 34, Created)
DT      01-OCT-1996 (rel. 34, Last sequence update)
DT      16-OCT-2001 (rel. 40, Last annotation update)
DE      Myosin II heavy chain.
GN      MYOJ.
OS      Dictyostellium discoideum (Slime mold).
OC      Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX      NCBI_TaxID=44689;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=AX3;
RX      MEDLINE=96215148; PubMed=8636147;
RA      Hammer J.A. III, Jung G.;
RT      "The sequence of the dictyostellium myo J heavy chain gene predicts a
RT      novel, dimeric, unconventional myosin with a heavy chain molecular
RT      mass of 258 kDa."
RL      J. Biol. Chem. 271:7120-7127(1996).
RN      (2)
RP      SEQUENCE OF 1-1021 FROM N.A.
RX      MEDLINE=97039016; PubMed=8884597;
RA      Peterson M.D., Urioste A.S., Titus M.A.;
RT      "Dictyostellium discoideum myof: a member of a broadly defined myosin
RT      V class or a class XI unconventional myosin?";
RL      J. Muscle Res. Cell Motil. 17:411-424(1996).
RN      (3)
RP      SEQUENCE OF 182-298 FROM N.A.
RX      MEDLINE=95023928; PubMed=7937787;
RA      Titus M.A., Kuspa A., Loomis W.F.;
RT      "Discovery of myosin genes by physical mapping in Dictyostellium.";
RL      Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
CC      - FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE
CC      ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC      - SUBUNIT: HOMODIMER.
CC      - SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC      - SIMILARITY: CONTAINS 3 IO DOMAINS.
CC      - SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL: U42409; AAA85186.1; -
DR      EMBL: L35322; AAA79858.1; -
DR      HSPR: P08799; 1MND.
DR      DictyDb: DP01095; myoj.
DR      InterPro: IPR002710; DIL.
DR      InterPro: IPR000048; IO_region.
DR      InterPro: IPR004009; Myosin_N.

```

DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00063; myosin\_head; 2.  
 DR Pfam: PF00612; IQ; 6.  
 DR Pfam: PF01843; DIL; 1.  
 DR Pfam: PF02736; Myosin\_N; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head; 2.  
 DR SMART: SM003376; DIL; 1.  
 DR SMART: SM00015; IQ; 3.  
 DR SMART: SM00242; MYSC; 1.  
 DR PROSITE: PS50096; IQ; 3.  
 KM Myosin: ATP-binding; Actin-binding; Calmodulin-binding; Repeat;  
 KM Coiled coil.  
 FT DOMAIN 1 809 MYOSIN HEAD-LIKE.  
 FT DOMAIN 824 851 IQ 1.  
 FT DOMAIN 872 901 IQ 2.  
 FT DOMAIN 943 972 IQ 3.  
 FT DOMAIN 973 1812 COILED COIL (POTENTIAL).  
 FT DOMAIN 1813 2245 TAIL.  
 FT NP\_BIND 174 181 ATP (POTENTIAL).  
 FT DOMAIN 669 749 ACTIN-BINDING.  
 FT CONFLICT 191 191 L -> F (IN REF. 2).  
 FT CONFLICT 284 284 A -> T (IN REF. 2).  
 FT CONFLICT 291 291 G -> R (IN REF. 2).  
 FT CONFLICT 332 347 MSGCFEIEGVSDEEH -> IEMFELKYRMKS (IN REF. 2).  
 FT CONFLICT 550 550 N -> K (IN REF. 2).  
 FT CONFLICT 865 866 HH -> QQ (IN REF. 2).  
 SO SEQUENCE 2245 AA; 258478 MW; 615E5EF1DA845BE CAC64;

Query Match 6.5%; Score 152; DB 1; Length 2245;  
 Best Local Similarity 23.6%; Pred. No. 0.11;  
 Matches 81; Conservative 57; Mismatches 143; Indels 62; Gaps 13;

QY 23 VYLLEKSAARPLRNENNAOTLRKLSQNPADALOEACLNKKKILSDSLYEFSK 82  
 DB 964 IQLAEKRSJRTVOEQNNKLOEKLEELQWRITSEAKRKQ--DEDKVKSDTTISLSS 1020  
 QY 83 TTRDIEL-ISWFAOPLDITLTLESANSL-EWLADSEK---HMDHLNPVLPEVTLK 135  
 DB 1021 NNHLELQLEIQKQVDELNSNSSLQLEKLEQTOQLDHSKLNKKLEKLSLSD 1080  
 QY 136 SDDDKGERQADAKKAFQVLGVDSSESSILYAPVLQPLVGEVTF---PDFQSAER- 190  
 DB 1081 QHDSIEKLOQFNETEQLOQFKQSESELSKSLKTTQ---QLDKNKQEFDLQSERD 1135  
 QY 191 -----KGEISQKSMLTVAOERFAIOFKMENAKRCYQLDLRLSLAVSTKHSLSOS 244  
 DB 1136 TDTNNOLEIQOLKK-ANSTLEEDYFSLGTRNLEKQVLELRDENQILKERLDSLQOQS 1194  
 QY 245 TNF--GFA-----KSLLTREVNALVHSGIKLAPKAER-----TVE 279  
 DB 1195 SDFQGAALKEQLEQVLQVQSEQLIKTSBKLGSEEEKKQINQLELTLDRKSLQIQ 1254  
 QY 280 QVAEASVS-----EGELPSHMDTK-----HIERIPMASEQQ 312  
 DB 1255 LQLTEQSNKIKKLKLEPDEQKQLOEQLEKRIKOSKOSVE 1297

RESULT 2  
 CC37\_YEAST  
 ID CC37\_YEAST STANDARD; PRT; 506 AA.  
 AC P06101; Q04132;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting subunit) (Cell division control protein 37).  
 GN CDC37 OR SMO1 OR YDR168W OR YD9489.03.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 NX NCBI\_TaxId=4932;

RN [1]  
 RP SEQUENCE OF 58-506 FROM N.A.  
 RX MEDLINE=86312926; PubMed=3018676;  
 RA Ferguson J., Ho J.-Y., Peterson T.A., Reed S.I.;  
 RT "Nucleotide sequence of the yeast cell division cycle start genes  
 RT CDC28, CDC36, CDC37, and CDC39, and a structural analysis of the  
 RT predicted products.";  
 RL Nucleic Acids Res. 14:6681-6697(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / AB972;  
 RA Oliver K., Harris D., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP INTERACTION WITH CDC28.  
 RX PubMed=7753858;  
 RA Gerber M.R., Farrell A., Deshaies R.J., Herskowitz I., Morgan D.O.;  
 RT "Cdc37 is required for association of the protein kinase Cdc28 with G1  
 RT and mitotic cyclins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:4651-4655(1995).  
 RN [4]  
 RP INTERACTION WITH STELL.  
 RX PubMed=10664467;  
 RA Abbas-Tekki T., Donze O., Picard D.;  
 RT "The molecular chaperone Cdc37 is required for Ste11 function and  
 RT pheromone-induced cell cycle arrest.";  
 RL FEBS Lett. 467:111-116(2000).  
 RN [5]  
 RP INTERACTION WITH CDC28 AND CAK1.  
 RX PubMed=10629030;  
 RA Farrell A., Morgan D.O.;  
 RT "Cdc37 promotes the stability of protein kinases Cdc28 and Cak1.";  
 RL Mol. Cell. Biol. 20:749-754(2000).  
 CC -1- FUNCTION: With Hsp90 it forms a complex that binds to several  
 CC kinases, resulting in stabilization and promotion of their  
 CC activity.  
 CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with CDC28, CAK1  
 CC and STELL.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE CDC37 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X04288; CAA27836.1; -;  
 DR EMBL: Z47813; CAA87799.1; -;  
 DR PIR: B26372; B26372.  
 DR SGD: S0002575; CDC37.  
 DR InterPro: IPR004918; Cdc37.  
 DR Pfam: PF03234; Cdc37; 1.  
 KW Chaperone; Cell division; Cell cycle.  
 FT CONFLICT 169 169 A -> D (IN REF. 1).  
 SO SEQUENCE 506 AA; 58385 MW; 0DF0C923158A2526 CRC64;

Query Match 5.4%; Score 126; DB 1; Length 506;  
 Best Local Similarity 19.1%; Pred. No. 0.64;  
 Matches 90; Conservative 71; Mismatches 161; Indels 148; Gaps 17;

QY 15 LSDPSICGVYLLKLEKSAARPLRNENNAOQLRLKLSQNPADALOEACLNKKKILSD 74  
 DB 13 LSDSDVEVPPNVDKKEF-----IKKK--QQ 36  
 QY 75 SLVEQFSKTRDIELISWFAOFLDITLTLESANSLLEWLADSEKHWDLNPVLEVTL 134  
 DB 37 SIHQREKRNQDILKNETQVDMYSHLNKRYDRI-----LSNPESSLTDLPAVTFPLNA 90  
 QY 135 KSDDDKGERQADAKKAFQVLGVDSSESSILYAPVLQPLVGEVTFPFQSAERKEI 194

Db 91 NEDMEKSGENVPRIATYNEKEDLFE-----OLA-----KDLDEKGD 131  
QY 195 SOLKSLTTTVAQERFAIQFKMENAKRCYTQDLRLSALVSTKCHSLGOSOTNGFAKSL 254  
Db 132 SKSPSLIRPAIKHAKIKSVTYEAK-----KLDELKREKNAHISSEDIHGFDSSEM 185  
QY 255 TRVENALVLSGI---KLAPKAEAKTVEQVASSVSE-----GEL 292  
Db 186 NKGKAGKALEATPSEALSAAES-NILNKLAKSSVPQFTIDKDDPMKAKTEERFGKI 244  
QY 293 PSMIDTKH-----IERIPMASEQ-----AQVSOHLHAG-----321  
Db 245 SINEYKSKQFLLEHLPIISEQKDALMKAEFYQLHGDKMTLQYIHOSELMAYIKEIY 304  
QY 322 -----NSELGNLNNNRDLAFHLREVSDFROSEPHSISFLLEKAIHWGLSLP 373  
Db 305 DMKKIPLYNMELSNVINM-----FPEKV--IFNKDKPMKESFL--RSVOKELHIO 353  
QY 374 E-----LLREMSQNGDALSTIFNAGLNHLDQVLLPEVSTPVGIESPO 419  
Db 354 KRSLIOGEEMDESNAEGVETI-----QLKSLDSTLEVLNLPFNSKDP 399  
RESULT 3  
REGA.DICDI STANDARD: PRT: 793 AA.  
AC 023917:  
DT 15-JUL-1998 (rel. 36, Created)  
DT 15-JUL-1998 (rel. 36, Last sequence update)  
DT 15-JUN-2002 (rel. 41, Last annotation update)  
DE 3',5'-cyclic-nucleotide phosphodiesterase rega (EC 3.1.4.17) (PDEase  
DE rega).  
GN REGA.  
OS Dictyostellum discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellum.  
ON NCBI\_TaxID=44689;  
RX (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-AX4:  
RA MEDLINE-97140317; PubMed-8986798;  
RA Shaulsky G., Escalante R., Loomis W.F.;  
RT "Developmental signal transduction pathways uncovered by genetic  
RT suppressors";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AX2:  
RA Thomason P.A., Traynor D., Cavet G., Chang W.T., Harwood A.J.,  
RA Kay R.R.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: MORPHOLOGICAL SUPPRESSOR OF TAGB.  
CC -1- CATALYTIC ACTIVITY: Nucleoside 3',5'-cyclic phosphate + H(2)O =  
CC nucleoside 5'-phosphate.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT LOW LEVELS IN VEGETATIVE CELLS  
CC AND AT HIGH LEVELS IN PRESPORE AND PRESTALK CELLS DURING  
CC DEVELOPMENT.  
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE  
CC FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: U60170; AAB03508.1; -  
CC EMBL: AU005398; CAA06513.1; -  
CC DictyDb: DD02055; rega.  
CC InterPro: IPR003607; ME\_Plase\_HDC.  
CC InterPro: IPR002073; PDEase.

DR InterPro: IPR001789; Response\_reg.  
DR Pfam: PF000723; response\_reg. 1.  
DR ProDom: PD000039; Response\_reg. 1.  
DR SMART: SM00471; HDC; 1.  
DR SMART: SM00448; REC; 1.  
DR PROSITE: PS00126; PDEASE\_1; 1.  
DR PROSITE: PS00110; RESPONSE\_REGULATORY; 1.  
KW Hydrolyase; CAMP; cGMP.  
FT DOMAIN 161 280  
FT DOMAIN 18 28  
FT DOMAIN 52 69  
FT DOMAIN 87 96  
FT DOMAIN 100 121  
FT DOMAIN 166 169  
FT DOMAIN 770 776  
SQ SEQUENCE 793 AA; 91175 MW; 6E065A6320F8C27 CRC64;  
Query Match 5.3%; Score 125; DB 1; Length 793;  
Best Local Similarity 19.0%; Pred. No. 1.3; Mismatches 184; Gaps 27;  
Matches 107; Conservative 88; Indels 184; Gaps 27;  
QY 1 MPLSKHQIQOLSKPLS-----DSDICGVYKLE---KSAERPLRNEF-----39  
Db 310 MELKEHEIEELTKKVSMSISKAMESPLVSTRNIEELKQSSMSHVESIEKLSI 369  
QY 40 -----NNAQTALRLKLSQNPASDERDALQELAKLWKIISDSLYQFSKTTDIELISW 92  
Db 370 LKELGSSNTRYRPSFEKLIKNDSDV-----PVKSFVSEFSTT-----408  
QY 93 FVAAGFLDTLTLSAANSLEMLADLSEKHWDLNPLYRVETLKSDD---GKERQADA 149  
Db 409 -----SRNST-----PRFQTYNRDRKEYKWE-----434  
QY 150 KVAFFQVGDSESSSTLYARVLQRLVGEVTFDFQSAE-RKEISQLSKSLTTTVAQE 208  
Db 435 -----FDVFKYSDD-----LMPLYDM-FENFQLRPIFKIPIEKLRFTMTVALY 480  
QY 209 R-----FALQFKME-----NAKRCYTQDLRLSALVSTKCHSLGOSOTNGFAK 251  
Db 481 RKNRRYNNFNHARDVQTUVTYFTLSFNAAOYLNLHLLFALLISCMDLNNPGFNTPQV 540  
QY 252 SLLTR-----VENALVHLSGIRKLAPKAEKTYE-----QEVASSVS---EGE 291  
Db 541 NAOTELSLLEYNDISVLENHNAMLT-FKILRNSCNILEGNEQYELKRSVQILLATD 599  
QY 292 LPSHMD-----TKHIERIPMASDAQTVSQHLHAGNLSELGNLNNNRDLAFHL-----L 341  
Db 600 MGNHFEHTNKFQHNLANLP--DRNKKEKQMLNLFILKGDLSNIAR--PWHLNFEWSL 655  
QY 342 REVSDYFROSEPHS-----PISFLEKA-----IRNGYL-----SLPELLREMA--380  
Db 656 RVSDDEFQOQSHYETIGYRPTPMDDTKTTRARIADDFIDFVASPLFQSMKAFKESQFL 715  
QY 381 -----SEQNGDALSTIFNAGLNHLDQVLLPEVSTPVGIESPOTPQAPVSDPRSYE 434  
Db 716 LKVISKNREWMQAYMEQKEGKCNDP---LQFMEDFTILVKS-KLKPIDEENRDKVSS 771  
QY 435 EHYVOTSQVDTQSKODQKQSS 456  
Db 772 SSSSSTAPLSTSSNNETSS 793  
RESULT 4  
MYH3\_HUMAN STANDARD: PRT: 1940 AA.  
ID MYH3\_HUMAN  
AC P11055; O15492;  
DT 01-JUL-1989 (rel. 11, Created)  
DT 01-JUL-1989 (rel. 11, Last sequence update)  
DT 16-OCT-2001 (rel. 40, Last annotation update)  
DE Myosin heavy chain, fast skeletal muscle, embryonic (muscle embryonic  
DE myosin heavy chain) (SMCE).  
GN MYH3.



| MYH3_RAT | STANDARD   | PRT | 1940 AA.                        |
|----------|--|-----|---------------------------------|
| AC       | P12847;  |     |                                 |
| DT       | 01-OCT-1989 (Rel. 12, Created)   |     |                                 |
| DT       | 01-OCT-1989 (Rel. 12, Last sequence update)  |     |                                 |
| DT       | 16-OCT-2001 (Rel. 40, Last annotation update)  |     |                                 |
| DE       | Myosin heavy chain, fast skeletal muscle, embryonic.   |     |                                 |
| CN       | MY3.   |     |                                 |
| OS       | Rattus norvegicus (Rat).   |     |                                 |
| OC       | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  |     |                                 |
| OC       | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  |     |                                 |
| OX       | NCBI_TaxID=10116;  |     |                                 |
| RN       | [1]  |     |                                 |
| RP       | SEQUENCE FROM N.A.   |     |                                 |
| RX       | MEDLINE=87060988; PubMed=3783701;  |     |                                 |
| RA       | Strehler E.E., Strehler-Page M.-A., Perliard J.C., Perlasamy M.,   |     |                                 |
| RA       | Nadal-Ginard B.;   |     |                                 |
| RT       | "Complete nucleotide and encoded amino acid sequence of a mammalian  |     |                                 |
| RT       | myosin heavy chain gene. Evidence against intron-dependent evolution   |     |                                 |
| RT       | of the rod.";  |     |                                 |
| RL       | J. Mol. Biol. 190:291-317(1986).   |     |                                 |
| CC       | - 1 FUNCTION: MUSCLE CONTRACTION.  |     |                                 |
| CC       | - 1 SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2   |     |                                 |
| CC       | HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  |     |                                 |
| CC       | AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).   |     |                                 |
| CC       | - 1 SUBCELLULAR LOCATION: Thick filaments of the myofibrils.   |     |                                 |
| CC       | - 1 DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  |     |                                 |
| CC       | CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,   |     |                                 |
| CC       | CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.   |     |                                 |
| CC       | - 1 PPM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  |     |                                 |
| CC       | ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.   |     |                                 |
| CC       | - 1 MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT   |     |                                 |
| CC       | MEROMYOSIN (LMN) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE   |     |                                 |
| CC       | SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED   |     |                                 |
| CC       | SUBFRAGMENT (S2).  |     |                                 |
| CC       | - 1 SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.   |     |                                 |
| CC       | - 1 SIMILARITY: CONTAINS 1 IQ DOMAIN.  |     |                                 |
| CC       | -----  |     |                                 |
| CC       | This SWISS-PROT entry is copyright. It is produced through a collaboration   |     |                                 |
| CC       | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |     |                                 |
| CC       | the European Bioinformatics Institute. There are no restrictions on its  |     |                                 |
| CC       | use by non-profit institutions as long as its content is in no way   |     |                                 |
| CC       | modified and this statement is not removed. Usage by and for commercial  |     |                                 |
| CC       | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |     |                                 |
| CC       | or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).  |     |                                 |
| CC       | -----  |     |                                 |
| DR       | EMBL: X04267; CAA27817.1; -  |     |                                 |
| DR       | PIR: A24922; A24922.   |     |                                 |
| DR       | HSSP: P13538; ZMTS.  |     |                                 |
| DR       | InterPro: IPR000048; IQ_region.  |     |                                 |
| DR       | InterPro: IPR004009; Myosin_N.   |     |                                 |
| DR       | InterPro: IPR002828; Myosin_tail.  |     |                                 |
| DR       | InterPro: IPR002017; Spectrin.   |     |                                 |
| DR       | InterPro: IPR001609; myosin_head.  |     |                                 |
| DR       | Pfam: PF00612; IQ; 2.  |     |                                 |
| DR       | Pfam: PF01576; Myosin_tail; 1.   |     |                                 |
| DR       | Pfam: PF02736; Myosin_N; 1.  |     |                                 |
| DR       | PRINTS: PR00193; MYOSINHEAVY.  |     |                                 |
| DR       | Prodom: PD000353; myosin_head; 1.  |     |                                 |
| DR       | SMART: SM00015; IQ; 1.   |     |                                 |
| DR       | SMART: SM00242; MYSC; 1.   |     |                                 |
| DR       | PROSITE: PS50096; IQ; 1.   |     |                                 |
| KW       | Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  |     |                                 |
| KW       | Calmodulin-binding; ATP-binding; Methylation; Alkylation;  |     |                                 |
| KW       | Multigene family.  |     |                                 |
| FT       | DOMAIN 1   | 781 | MYOSIN HEAD-LIKE.               |
| FT       | DOMAIN 1   | 811 | IQ.                             |
| FT       | DOMAIN 1   | 840 | COILED COIL (POTENTIAL).        |
| FT       | NP_BIND  | 179 | ATP (POTENTIAL).                |
| FT       | DOMAIN   | 179 | ACTIN-BINDING.                  |
| FT       | DOMAIN   | 758 | ACTIN-BINDING.                  |
| FT       | MOD_RES  | 130 | METHYLATION (TRI-) (POTENTIAL). |

| FT  | MOD_RES   | 696   | 696                           | ALKYLATION (SH-1)       |
|---|---|---|-------------------------------|-------------------------|
| FT  | MOD_RES   | 706   | 706                           | ALKYLATION (SH-2)       |
| 5Q  | SEQUENCE  | 1940 AA;  | 223857 MW;                    | B5D546A596B5A696 CRC64; |
| Query Match   |   |   |                               |                         |
| Best local similarity 20.9%; Pred. No. 4.9;                       |   |   |                               |                         |
| Matches 101; Conservative 90; Mismatches 194; Indels 99; Gaps 21. |   |   |                               |                         |
| QY  | 7   | QIEQLSKPLSDSDSTICGVYKLEKSAFR---                               | PLRNEFVNAQTALRKLSONPSADERDALQ | 62                      |
| Db  | 919   | KIKVEYTERAEDEEETINAEITAKRKLDEBCELSKKDIDOLETLAKVEKEKHATE-----  |                               | 973                     |
| QY  | 63  | EACLNKKKILSDSL---EQFSKTRDIELLSWFAAQFLD--                      | TTLESAAASLEMLADL              | 117                     |
| Db  | 974   | ---NKVNLITTEELGLDDETLAKTLRREKALD--EAAHOQTLDDQLAEDBKVNSLSKLKS  |                               | 1027                    |
| QY  | 118   | SEKHMDHNPVLPVE-TLKSDDDKCKERQADAVK--KAFFQYDPSSESSILVAPVLQ      |                               | 174                     |
| Db  | 1028  | LEQVYDDLESSLEQKKLRLVDERKKRKLLEGDLKLAQESIDLELNDKQO-----        |                               | 1076                    |
| QY  | 175   | PLVGEVTFEEDFQSAERKGEISQLKSMLTITTVAAQER---                     | FAIOFKEMNAKRCVTOL-DRLS        | 230                     |
| Db  | 1077  | -----LDEKLKKKKDEYEQLOS-----KVDEQDTLSLQDKKIKELQARIEELEEEIE     |                               | 1124                    |
| QY  | 231   | ALVSTKCHSLGSQSTNEGFANSLTRVENA-LVHLSGIKLAPKAEK--               | TVQOEVAESSV                   | 287                     |
| Db  | 1125  | AERATRAKTEKORSYARELEBELSERLEBAGCVSTQTOIELNKKREAEFLKLRLDEEATL  |                               | 1184                    |
| QY  | 288   | SEGELPSIMDPRKHIERIMASQAOQTVOHLAGNLSELGNLNNNRDL----            | AFHLIR                        | 342                     |
| Db  | 1185  | QHEATVAFLRKKNAD--SAAELAE-----QIDNLQRYKQKLEKESSEFKL--          |                               | 1227                    |
| QY  | 343   | EVSDYFROSEPHSPISFLKEKALRMGYLSLPELREMKSEONGDALSTIFNAGLHLDQ     |                               | 402                     |
| Db  | 1228  | EIDDLSSVESVSKKANLEKICR-----TLEDQlse-----                      | ARGKNIEEQ                     | 1268                    |
| QY  | 403   | VLLPEVSPPTVGIESPQTPQAKPSVSDPSVEHVSQTSPPVDQSKODKQPO---         | SSAT                          | 458                     |
| Db  | 1269  | RLSELTLTQKSLQF--EAGELSLRQLEEKESIVSLSRSKQAPFTQOIEELKRLQLEEBNKA |                               | 1327                    |
| QY  | 459   | SALS  | 462                           |                         |
| Db  | 1328  | NALA  | 1331                          |                         |
| RESULT 6  |   |   |                               |                         |
| MSP_SCHEMA  |   |   |                               |                         |
| ID  | MSP_SCHEMA  | STANDARD:   | PRT:                          | 866 AA.                 |
| AC  | P06198;   |   |                               |                         |
| DT  | 01-JAN-1988 (Rel. 06, Created)  |   |                               |                         |
| DT  | 01-JUN-1984 (Rel. 29, Last sequence update)                           |   |                               |                         |
| DT  | 16-OCT-2001 (Rel. 40, Last annotation update)                         |   |                               |                         |
| DE  | Paramyosin.   |   |                               |                         |
| OC  | Schistosoma mansoni (Blood fluke).                                    |   |                               |                         |
| OC  | Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; |   |                               |                         |
| OC  | Schistosomatoidea; Schistosomatidae; Schistosoma.                     |   |                               |                         |
| OX  | NCBI_TaxID=6183;  |   |                               |                         |
| RN  | [1]   |   |                               |                         |
| RP  | SEQUENCE FROM N.A.  |   |                               |                         |
| RA  | MEDLINE=91270282; PubMed=2052029;                                     |   |                               |                         |
| RA  | Lajelette J.P., Landa A., Arcoo L., Williams K., Davis A.E.,          |   |                               |                         |
| RA  | Shemmaker C.B.;   |   |                               |                         |
| RT  | "Paramyosin is the Schistosoma mansoni (Trematoda) homologue of       |   |                               |                         |
| RT  | antigen B from Taenia solium (Cestoda)."                              |   |                               |                         |
| RL  | Mol. Biochem. Parasitol. 44:287-296(1991).                            |   |                               |                         |
| RN  | [2]   |   |                               |                         |
| RP  | SEQUENCE OF 303-742 FROM N.A.   |   |                               |                         |
| RA  | MEDLINE=87018840; PubMed=3094144;                                     |   |                               |                         |
| RA  | Lanar D.E., Pearce E.J., James S.L., Sher A.;                         |   |                               |                         |
| RT  | "Identification of paramyosin as schistosome antigen recognized by    |   |                               |                         |
| RT  | intradermally vaccinated mice."                                       |   |                               |                         |
| RL  | Science 234:593-596(1986).  |   |                               |                         |
| CC  | -I- FUNCTION: PARAMYOSIN IS A MAJOR STRUCTURAL COMPONENT OF           |   |                               |                         |

CC MANY THICK FILAMENTS ISOLATED FROM INVERTEBRATE MUSCLES.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -1- SIMILARITY: HIGH, TO MYOSIN HEAVY CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M35499; AAA29915.1; -;  
 DR EMBL: M4163; AAA29914.1; -;  
 DR PIR: A25993; A25993.  
 DR InterPro: IPR002928; Myosin tail.  
 DR Pfam: PF01576; Myosin\_tail; 1.  
 KW Coiled coil; Muscle protein; Thick filament; Myosin.  
 FT DOMAIN 1 22 NONHELICAL REGION (POTENTIAL).  
 FT DOMAIN 23 839 COILED COIL (POTENTIAL).  
 FT DOMAIN 840 866 NONHELICAL REGION (POTENTIAL).  
 FT DISULFID 750 750 INTERCHAIN (POTENTIAL).  
 FT CONFLICT 423 428 DOYKEL -> GSSRT (IN REF. 2).  
 FT CONFLICT 430 431 SS -> KL (IN REF. 2).  
 FT CONFLICT 637 637 V -> L (IN REF. 2).  
 FT CONFLICT 639 639 T -> I (IN REF. 2).  
 FT CONFLICT 691 691 G -> E (IN REF. 2).  
 FT CONFLICT 720 720 S -> F (IN REF. 2).  
 SQ SEQUENCE 866 AA; 100387 MW; 42FA56E78176AE0 CRC64;  
 Query Match 5.3%; Score 124; DB 1; Length 866;  
 Best Local Similarity 21.3%; Pred. No. 1.8;  
 Matches 103; Conservative 82; Mismatches 182; Indels 116; Gaps 21;

QY 7 QIBLSKPLSDSICGYVLEKSAFRLRNFVNAQTALFKLSQNSPADRDALQEQCL 66  
 DB 122 EVELNQKQK-----KAEDKSHLMEVDNVLQDGLAKAKQSAESKLEGLDSOL 172  
 QY 67 NKMKILSDLYEOPSK-----TTTRDIEL-TSMFVAQFLDTEESANSLEWLAD 116  
 DB 173 NRKLSLTDQRLQTELENNAKSRILTSEFELHINQDEAQL---NYSKAKSSLEQYD 229  
 QY 117 LSEKHWHLNVLVEVETLAKSDDKGKEREQADAKVAFQVLGD-----SEESSIL 167  
 DB 230 -----DLKRSLLDEAKNFFNLQAOLTS-LQMDYDNLQAKYDESEESASNL 273  
 QY 168 YAPVILQPLVGEVTFEPDQSAERKEISQKSMLTTFYAOERFALIOFMENAKKCVQLD 227  
 DB 274 RS-----QVSKFNDAIALAKSKFE--RELMSKTEEFEMKRFKTFMR-----ITELE 317  
 QY 228 RLIALVSTKCHSLGSGSTNFGF-AKSLTLRVENALVHLGKILAKPAKAEKTVDEQVAVSS 286  
 DB 318 DTAERERKAVSLKELKTKLLEIKDLQSEIES--LSLENSLEIRRAKA-----AESL 368  
 QY 287 VSEBELSHMDTKIETIRPMASEQAQTVSOHLHAGNL-SELGINLN-MNRDLAFHLREV 344  
 DB 369 AS--DLQRRVDLELTIEVNTLSQNSQSENLRLKSLVNDLTDRKNLLEIRE-----NRQM 421  
 QY 345 SDYFRQSPHSPISFLLEKAIKMGYLSPEL--LREMMSENGALSTIFNA----- 394  
 DB 422 NDQYKE-----LKSSLRDANRRRTDLEALRSOLEARDNLASLHDAEALHDM 471  
 QY 395 -----AGLNHLDOVLLPEVSTPTVGIESPOTQAPKPSVDPSVEHVSQTSVPVDTOS 447  
 DB 472 QKYASQAALNHLKSEMGRLRERDEELSLRK-----STTTIEELTYTITMEVKY 524  
 QY 448 KOD 450  
 DB 525 KSE 527

RESULT 7

DP01\_RICPR  
 ID DP01\_RICPR STANDARD; PRT; 867 AA.  
 AC 005949;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA polymerase I (PC 2.7.7.7) (POL I).  
 GN POLA OR RP776.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.  
 CC NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Madrid E;  
 RX MEDLINE=97419517; PubMed=9274032;  
 RA Andersson J.O., Andersson S.G.E.;  
 RT "genomic rearrangements during evolution of the obligate  
 RT intracellular parasite Rickettsia prowazekii as inferred from an  
 RT analysis of 52015 bp nucleotide sequence";  
 RL Microbiology 143:2783-2795(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Madrid E;  
 RX MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria";  
 RL Nature 396:133-140(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Madrid E; and B;  
 RX MEDLINE=99416441; PubMed=10486973;  
 RA Andersson J.O., Andersson S.G.E.;  
 RT "genome degradation is an ongoing process in Rickettsia";  
 RL Mol. Biol. Evol. 16:1178-1191(1999).  
 CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE  
 CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate  
 CC + [DNA](N).  
 CC -1- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y11784; CAAT72476.1; -;  
 DR EMBL: AJ235273; CAAT5203.1; -;  
 DR EMBL: AJ238755; CAB56085.1; -;  
 DR EMBL: AJ238756; CAB56089.1; -;  
 DR HSSP: P00582; 1KFS.  
 DR InterPro: IPR002421; 5\_3\_exonuclease.  
 DR InterPro: IPR001098; DNA\_pol.  
 DR InterPro: IPR002298; DNA\_pol.  
 DR InterPro: IPR000513; Exo\_N.I.  
 DR InterPro: IPR003584; HHH\_2.  
 DR Pfam: PF004476; DNA\_pol\_A; 1.  
 DR Pfam: PF01367; 5\_3\_exonuclease; 1.  
 DR Pfam: PF02739; 5\_3\_exonuc; N; 1.  
 DR PRINTS: PR00868; DNAPOLI.  
 DR SMART: SM00475; 53EXOC; 1.  
 DR SMART: SM00279; Hhh2; 1.  
 DR SMART: SM00482; POLAC; 1.  
 DR TIGRfams: TIGR00593; pola; 1.

```

DR PROSITE PS00447: DNA POLYMERASE A; 1.
KW Transferrase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW Hydrolase; Exonuclease; DNA-binding; Complete proteome.
FT DOMAIN 1 283 EXONUCLEASE.
SQ SEQUENCE 867 AA; 98651 MW; 557194B38FB69BA2 CRC64;

Query Match 5.3%; Score 123.5; DB 1; Length 867;
Best Local Similarity 21.2%; Pred. No. 1.9; Mismatches 172; Indels 103; Gaps 17.
Matches 90; Conservative 60;

QY 18 DSICGVYLKLEASAPRPLNEFNAQTALRKISQNPASADERALQDACLNKKILSDSLY 77
DB 186 DNIPEVPISGPKTA--SLITKFGSEVNFNSLDQISSIKQKRTLONA----- 231
QY 78 EGFSTRTDIELISMFVAQFLDPTT--LESANSLLEWLADISEKHMDLNPVLYETLK 135
DB 232 -----REALISMKLIG--LDSENVLDLFDLNNLKMSPPRSKK---LTGFLQEGPK 277
QY 136 SDDDKGREQADAKYKAFQQL--VQDSESSILYAPVQLPLVGEVTFDFQSAERKGEI 194
DB 278 S-----LYKRVENLFDIKINDHE-----IYDNKVTEAK-EI 308
QY 195 SOLKMLTTTVAQER-----FAIDFKEMNAKRCYTQDLRSALVSTKCHSLGSGSTNG 248
DB 309 SNASLANPAKAEIRIGITGLYLDQKGEN--RALLISLQNSQYIITKITTNNNNIKNNNDW 367
QY 249 FAKSLLTREVEN-----ALVHLGSIKAPKAEKATVEQ--EVAESSVSEGEPLSHMDK 299
DB 368 FSHILNLNLTKNSIKRTIYSLKHLKLFYANOSHQTALIEDLEMQALYSAGLYOKNLPTK 427
QY 300 HIERIPMASEQAQTVSQ--HLHAGNLSELGNLNNNRDLAFHLREVSDYFROSEPHSPI 357
DB 428 TLTKNININESARIYINFIYKQTLLEL-----QKNKFAFLYREID-----LPT 472
QY 358 SFLLEKAIKRWGLSLPELLREMSQNGDAL-----STIRPAGLNHLDOYLPE 407
DB 473 CGLDKMEKVGKIVDANYLNRLSDFGTEILKIEEIEFALSGTKFNIGSQKOLGELLFK 532
QY 408 VSTPT 412
DB 533 MQLPS 537

RESULT 8
MYH7_HUMAN STANDARD: PRT: 1935 AA.
AC P12883; Q14904; Q16579.
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle beta isoform (MYHC-beta).
GN MYH7 OR MYHCB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9105634; PubMed=2249844;
RX Janelicke T., Diederich K.W., Haas W., Schleich J., Lichter P.,
RX Pfordt M., Bach A., Vobberg H.P.;
RX "The complete sequence of the human beta-myosin heavy chain gene and
RX a comparative analysis of its product.";
RX Genomics 8:194-206(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90301496; PubMed=2362820;
RX Liew C.-C., Sole M.J., Yamauchi-Takahara K., Kellam B.,
RX Anderson D.H., Ian L., Liew J.;
RX "Complete sequence and organization of the human cardiac beta-myosin
RX heavy chain gene.";
RX Nucleic Acids Res. 18:3647-3651(1990).
RN [3]

```

RP SEQUENCE OF 1-115 FROM N.A.  
 RX MEDLINE=69264452; PubMed=2726733;  
 RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;  
 RT "Characterization of human cardiac myosin heavy chain genes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).  
 RN [14]  
 RP ERRATUM.  
 RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).  
 RN [15]  
 RP SEQUENCE OF 1310-1935 FROM N.A.  
 RX MEDLINE=86176778; PubMed=24212254;  
 RA Saez U., Lelwand L.A.;  
 RT "Characterization of diverse forms of myosin heavy chain expressed in  
 RL adult human skeletal muscle";  
 RN Nucleic Acids Res. 14:2951-2969(1986).  
 RP REVISIONS.  
 RA Lelwand L.A.;  
 RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.  
 RN [17]  
 RP SEQUENCE OF 1410-1935 FROM N.A.  
 RX MEDLINE=88299163; PubMed=2965919;  
 RA Kurabayashi M., Tschimochi H., Komuro I., Takaku F., Yazaki Y.;  
 RT "Molecular cloning and characterization of human cardiac alpha- and  
 RL beta-form myosin heavy chain complementary DNA clones. Regulation of  
 RT expression during development and pressure overload in human  
 RL actin";  
 RN J. Clin. Invest. 82:524-531(1988).  
 RP [8]  
 RP SEQUENCE OF 785-1935 FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=90235862; PubMed=1691980;  
 RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goede H.W.,  
 RN Arnold H.H.;  
 RT "Identification of three developmentally controlled isoforms of human  
 RL myosin heavy chains";  
 RN Eur. J. Biochem. 189:55-65(1990).  
 RP [19]  
 RP SEQUENCE OF 1393-1935 FROM N.A.  
 RX MEDLINE=87192738; PubMed=8032765;  
 RA Jandreski M.A., Liew C.-C.;  
 RT "Construction of a human ventricular cDNA library and  
 RL characterization of a beta myosin heavy chain cDNA clone";  
 RN Hum. Genet. 76:47-53(1987).  
 RP [10]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=96039076; PubMed=8533830;  
 RA Arai S., Matsuoaka R., Hirayama K., Sukurai H., Tamura M., Ozawa T.,  
 RA Kikura M., Imamura S.-I., Fututani Y., Joh-O K., Kawana M., Takao A.,  
 RN Hosoda S., Momma K.;  
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in  
 RL hypertrophic cardiomyopathy";  
 RN Am. J. Med. Genet. 58:267-276(1995).  
 RP [11]  
 RP VARIANTS CMH1 GLU-256 AND ARG-741.  
 RX MEDLINE=93248216; PubMed=8483915;  
 RA Fananapazir L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.;  
 RT "Missense mutations in the beta-myosin heavy-chain gene cause central  
 RL core disease in hypertrophic cardiomyopathy";  
 RN Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).  
 RP [12]  
 RP VARIANT CMH1 GLN-403.  
 RX MEDLINE=90367131; PubMed=1975517;  
 RA Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,  
 RA McKenna W., Selman C.E., Seidman J.G.;  
 RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta  
 RL cardiac myosin heavy chain gene missense mutation";  
 RN Cell 62:999-1006(1990).  
 RP [13]  
 RP VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.  
 RX MEDLINE=92204193; PubMed=1552912;  
 RA Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W.,

RA Seidmann C.E., Seidmann J.G.:  
RT "Characteristics and prognostic implications of myosin missense  
RT mutations in familial hypertrophic cardiomyopathy.";  
RL New Engl. J. Med. 326:1108-1114(1992).  
RN [14]  
RP VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.  
RX MEDLINE-94070863; PubMed-8250038;  
RA Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A.,  
RT McKenna W., Seidman J.G., Seidman C.E.:  
RT "Independent origin of identical beta cardiac myosin heavy-chain  
RT mutations in hypertrophic cardiomyopathy.";  
RL Am. J. Hum. Genet. 53:1180-1185(1993).  
RN [15]  
RP VARIANTS CMH1 GLN-403 AND VAL-908.  
RX MEDLINE-92346810; PubMed-1638703;  
RA Epstein N.D., Cohn G.M., Cyran F., Fananapazir L.:  
RT "Differences in clinical expression of hypertrophic cardiomyopathy  
RT associated with two distinct mutations in the beta-myosin heavy chain  
RT gene. A 908Leu->Val mutation and a 403Arg->Ser mutation.";  
RL Circulation 86:345-352(1992).  
RN [16]  
RP VARIANTS CMH1 LEU-403 AND TRP-403.  
RX MEDLINE-94075629; PubMed-8254035;  
RA Dausse E., Komajda M., Fellet L., Dubourg O., Dufour C., Carrier L.,  
RT Wisniewsky C., Bercovici J., Hengstenberg C., Al-Mahdawi S.:  
RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and  
RT identification of a hot spot for mutations in the beta-myosin heavy  
RT chain gene.";  
RL J. Clin. Invest. 92:2807-2813(1993).  
RN [17]  
RP VARIANTS CMH1 TRP-403.  
RX MEDLINE-94093568; PubMed-8268932;  
RA Moolman J.C., Brink P.A., Corfield V.A.:  
RT "Identification of a new missense mutation at Arg403, a Cpg mutation  
RT hotspot, in exon 13 of the beta-myosin heavy chain gene in  
RT hypertrophic cardiomyopathy.";  
RL Hum. Mol. Genet. 2:1731-1732(1993).  
RN [18]  
RP VARIANTS CMH1 ASN-615.  
RX MEDLINE-93038688; PubMed-1417858;  
RA Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.:  
RT "Novel missense mutation in cardiac beta myosin heavy chain gene  
RT found in a Japanese patient with hypertrophic cardiomyopathy.";  
RL Biochem. Biophys. Res. Commun. 188:379-387(1992).  
RN [19]  
RP VARIANTS CMH1 GLY-778.  
RX MEDLINE-93343938; PubMed-8343162;  
RA Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.:  
RT "A missense mutation of cardiac beta-myosin heavy chain gene linked  
RT to familial hypertrophic cardiomyopathy in affected Japanese  
RT families.";  
RL Biochem. Biophys. Res. Commun. 194:791-798(1993).  
RN [20]  
RP VARIANTS CMH1 VAL-908.  
RX MEDLINE-93168465; PubMed-8435239;  
RA Al-Mahdawi S., Chamberlain S., Cleland J., Mihoyannopoulos P.,  
RT Gilligan D., French J., Choudhury L., Williamson R., Oakley C.:  
RT "Identification of a mutation in the beta cardiac myosin heavy chain  
RT gene in a family with hypertrophic cardiomyopathy.";  
RL Br. Heart J. 69:136-141(1993).  
RN [21]  
RP VARIANTS CMH1 TRP-719.  
RX MEDLINE-95179132; PubMed-7874131;  
RA Greve G., Bachinski L., Friedman D.L., Czernezewicz G., Anan R.,  
RT Towbin J., Seidman C.E., Roberts R.:  
RT "Isolation of a de novo mutant myocardial beta MHC protein in a  
RT pedigree with hypertrophic cardiomyopathy.";  
RL Hum. Mol. Genet. 3:2073-2075(1994).  
RN [22]  
RP VARIANTS CMH1 CYS-513; ARG-716 AND TRP-719.  
RX MEDLINE-94110336; PubMed-8282798;  
RA Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S.,  
RT Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A.,

RA Spirito P., Roberts R., Seidman J.G., Seidman C.E.:  
RT "Prognostic implications of novel beta cardiac myosin heavy chain gene  
RT mutations that cause familial hypertrophic cardiomyopathy.";  
RL J. Clin. Invest. 93:280-285(1994).  
RN [23]  
RP VARIANTS CMH1 THR-797.  
RX MEDLINE-96047159; PubMed-7581410;  
RA Moolman J.C., Brink P.A., Corfield V.A.:  
RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-  
RT myosin heavy chain gene in hypertrophic cardiomyopathy.";  
RL Hum. Mutat. 6:197-198(1995).  
RN [24]  
RP VARIANTS CMH1 CYS-453.  
RX MEDLINE-96209901; PubMed-8655135;  
RA Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C.,  
RT Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.:  
RT "Malignant familial hypertrophic cardiomyopathy in a family with a  
RT 453Arg->Cys mutation in the beta-myosin heavy chain gene:  
RT coexistence of sudden death and end-stage heart failure.";  
RL Hum. Genet. 97:585-590(1996).  
RN [25]  
RP VARIANTS CMH1 THR-349 AND TRP-719.  
RX MEDLINE-98204402; PubMed-9544842;  
RA Jeschke B., Uhl K., Weist B., Schroder D., Metlinger T.,  
RT Dohleman C., Vosberg H.-P.:  
RT "A high risk phenotype of hypertrophic cardiomyopathy associated with  
RT Query Match 5.2%; Score 123; DB 1; Length 1935;  
Best Local Similarity 22.1%; Pred. No. 6;  
Matches 94; Conservative 63; Mismatches 140; Indels 128; Gaps 19;  
QY 7 QIEQLSKPL-----SDSICGVYK--LEKSAFRLRNEF-----NVAQTALR 47  
DB 1385 ELERAKKKLAORLOAEAEVAVNAKCSLEKTKHR-LQNEIEDLVWVERNAALD 1443  
QY 48 KLSKN-----PSADERDALQKAC-----LKKMKTLSD 74  
DB 1444 KQKRNFKLILAEWKQKYEESQSELESSQKARSSTLEFLKNAEESLEHLETFKRNK 1503  
QY 75 SLVYQFS-----KTRDIELISFVAQFL-LDTTLESANSLFWLADLSEKHW 123  
DB 1504 NLQGEISDLNEOLQSSKTTHELEKVKQLEAEKMELOSALKEAELE-----HEEG 1557  
QY 124 HLNPLVPLETFLKSD-----DDKGRERBA--DAKYAFQVLQVDSSESSILYAPVLQ 175  
DB 1558 ILRQLEFQNIKAIERKLEKDEEMQAKRNHLRAVDLSQTSIDATRNRNEALRVK 1617  
QY 176 LVGEVTFDFDQ-----SARKGEISQLKSMLTITVAQ-----ERRAT---- 212  
DB 1618 MEGDLNEMETIOLSHANMAAEQKQVKSLOSLSLKDPTIOLDDAVRANDLKEMIAI 1677  
QY 213 ----QFKMENAKRCVPTOLDRSLA-----VSTKCHSLGSSQTNFGFAK-----SL 253  
DB 1678 NNLQAELEELRAVDEQTESRKLAEOLEIETSRVOLLHSQNTSLNQKKKDDADLSOL 1737  
QY 254 LTRVENALVHLGKILKPAKEAKTVEQVAESSVSEGELEPSSHDTK-HIERIPMAEQAQ 312  
DB 1738 QTEVEEA-----VQECRNMEK-AKKAITDAAMMAELKKEQDTSNHLRMRKKNMQ 1790  
QY 313 TVSQH 317  
DB 1791 KDLOH 1795  
RESULT 9  
MYH7\_RAT  
ID MYH7\_RAT STANDARD; PRT; 1935 AA.  
AC P02564;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin heavy chain, cardiac muscle beta isoform (MYHC-beta).  
GN MYH7.



OS Onchocerca volvulus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
OC Onchocercidae; Onchocerca.  
OX NCBI\_TaxID=6282;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95287898; PubMed=7770081;  
RA Triteeraprab S., Richle T.L., Tuan R.S., Shepley K.J., Dimman J.D.,  
RA Neubert T.A., Scott A.L.;  
RT "Molecular cloning of a gene expressed during early embryonic  
RT development in Onchocerca volvulus.";  
RL Mol. Biochem. Parasitol. 69:161-171(1995).  
RN  
RP SEQUENCE OF 733-866 FROM N.A.  
RX MEDLINE=89127417; PubMed=2464764;  
RA Donaldson J.E., Duke B.O.L., Moser D., Zeng W., Erondou N.E.,  
RA Lucius R., Renz A., Karam M., Flores G.Z.;  
RT "Construction of Onchocerca volvulus cDNA libraries and partial  
RT characterization of the cDNA for a major antigen.";  
RL Mol. Biochem. Parasitol. 31:241-250(1988).  
CC -1- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U12681; AAA80009.1; -;  
DR EMBL: J03995; AAA29412.1; -;  
DR PIR: A54513; A54513.  
KW Antigen; Coiled coil.  
FT DOMAIN 74 120 COILED COIL (POTENTIAL).  
FT DOMAIN 151 251 COILED COIL (POTENTIAL).  
FT DOMAIN 327 384 COILED COIL (POTENTIAL).  
FT DOMAIN 417 1879 COILED COIL (POTENTIAL).  
SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;  
Query Match 5.2%; Score 123; DB 1; Length 2022;  
Best Local Similarity 22.9%; Pred. No. 6.4; Indels 82; Gaps 16;  
Matches 85; Conservative 64; Mismatches 140;  
QY 35 LRNEFVAQTALRKLSQNPASDERDALQEAOLKMKILSDLSLYEOPSKTRDIELISFEV 94  
DB 971 LRKLNDAERAMADL-QN-----RDSILERENDMKREKSDALNNEIDLR-LRD-ELLSVRR 1022  
QY 95 AAOFL--DTTESAA-NSLEWLADISEKHMDHLNVLFEVETLKSDDDKGKERQADAK 150  
DB 1023 DAEKEINRYTDTQTAARNEIKLPTNNMKSQLN-----AAEDK 1063  
QY 151 VKAFVOLGVSESSILYAVLQLPLVGEVTFDFEOSAEKRGESLSOLKMLTTVAERF 210  
DB 1064 INSLNKYITQONKI-----RDLTGEVHHLEGLKDKGNVANLESDPTT---RE 1111  
QY 211 AIOFKMENARCYOTDLRLSALVSTKCHSLGOSTNFGFAKSLTTRVENALVHLGSIKLA 270  
DB 1112 RIHLGQNSLQTELKIK-----GDIDSLFG-----ENM-LTKAKES 1150  
QY 271 PKAEAKVEDEVAESSSEELPSHMDTKHIERIPMASEOQVYVSOHLHAGNLSGLGNL 330  
DB 1151 NEAEIRLKKOKLORSTENAKKYSALDKLRPEYDRL-----QMLYREKIKQAEMLT 1201  
QY 331 NMNRDLAFHL-----LREYSDFROSE-PHSPISFLLEKAIKRGYSLDELLREMMSEQ 383  
DB 1202 QAVODLESRLNOSRRRLRDATKLIASEGDRNALRSEVERKLOHVEQPMRQLLKK--TDE 1259  
QY 384 NGDALSTIFNA 394  
DB 1260 YQAALSDLVNA 1270

RESULT 11  
TEGU\_HSVSA STANDARD; PRT; 2469 AA.  
ID TEGU\_HSVSA  
AC 001056;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DE 01-APR-1993 (Rel. 25, Last annotation update)  
DT Probable large tegument protein.  
GN 64 OR ERF2.  
OS Herpesvirus saimiri (strain 11).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Camaherpesvirinae; Rhadinovirus.  
OX NCBI\_TaxID=10383;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92230228; PubMed=1314457;  
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;  
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,  
RA Newman C., Wiltmann S., Craxton M.A., Coleman H., Fleckenstein B.,  
RA Honess R.W.;  
RT "Primary structure of the herpesvirus saimiri genome.";  
RL J. Virol. 66:5047-5058(1992).  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92230228; PubMed=1314457;  
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;  
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of  
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic  
RT organization between HVS and Epstein-Barr virus.";  
RL Virology 188:286-310(1992).  
CC  
CC -1- FUNCTION: TEGUMENT PROTEIN.  
CC  
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,  
CC EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X64346; CAA55687.1; -;  
DR EMBL: M66409; AAA46140.1; -;  
DR PIR: H36812; H36812.  
SQ SEQUENCE 2469 AA; 280165 MW; D2B4B8DC08644CDB CRC64;  
Query Match 5.2%; Score 122.5; DB 1; Length 2469;  
Best Local Similarity 20.5%; Pred. No. 9; Indels 221; Gaps 27;  
Matches 121; Conservative 77; Mismatches 172;  
QY 40 NVAQOTALRKLSQNPASDERDALQ-----EACLKMKILSDLSLYEOPSKTRR 85  
DB 694 NVLTDLHHTKKQKQTOISPAPDFTYVKNQSTQLQGLQTCVDTLNT-DKFF--ISMTVQ 749  
QY 86 DIELISFEVAOFLDITTESAANSLEWLADISEKHMD-HLNPVLFEVETLKSDDDKGKE 143  
DB 750 QLSYIGVEV-----AELSHSQWNEFKADPVLPKLITL--DDIKKE 786  
QY 144 RQADAKVK-----AFQO-LVGDSESSILYAVLQL-----PLVGEVTFDFEOS 187  
DB 787 IQOVTTKQKNEEPLSKLILAVQVOTLENAKOSDTLSITPDIHYTTKAGTLGGERENOKFES 846  
QY 188 AE---RKGEISQ--LKSMLTTVAOEFATOFMENAKRCVOTDLRLSALVSTKCHSLGS 242  
DB 847 LKNTVQKLSLSEEFKLTIDST-----LENVQ--LDIQEISDILQSNQYIHQS 893  
QY 243 QSTNFGFAKSLTRVENALVHLSGIR-----LAPK--AAKATVEO-----EVAESS 286  
DB 894 ETIKQAEFFDKSNITINNIIQILNOKKTYTPTOPMLIAVKRFLSEAKFRESENTICEIISTL 953  
QY 287 VSEGEPLSHMDTKHIERIPMASQOQVYVSOHLHAGNLSGLGNLNMNRDLAFHL--LREV 344  
DB 954 VSLGSLISKSTT--VEALKDALKSIDTLKERTLAVDPKRLRELYNVRKLOKQKLTLLQ 1011

QY 345 SDY-----FROSEPHSPISFLEKAIKRWGYLSL-----PELLREM 379  
DB 1012 QEDDNKMEVDSPVPTPSRDVKTFFIONAPSMKAKQYAKKALKKOIOAMEDVDPESEYIED 1071  
QY 380 MSQONGDAL-----STIFNAG-----396  
DB 1072 NIKANGOKAMQKIQSAFODLNFSILIPDWSLAKETRPKSTLFVIGPILKFEVEVL 1131  
QY 397 -----LN-HLDQVLLPEVST--PVGIESP 418  
DB 1132 ESKVNLKEAKLKLLPNGVPVFTPEPKEDMIHYESNVNHFILKTINLPVSTVANHIGHEL 1191  
QY 419 QTFQAKPSVSDPR-----SVEEHVSQTS-----PVDTSKQDO 451  
DB 1192 LLSQALNSKTLPEAVYGTSLDQHAAKFSKFKTLLEATWHQVDTKRIKIDE 1242

RESULT 12  
MYH7\_MESAU  
ID MYH7\_MESAU STANDARD: PRT: 1934 AA.  
AC P13540: 060540:  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).  
GN MYH7.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OX Mesocricetus.  
NCBI\_TaxID=10036;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=F1B; TISSUE=Liver;  
RX MEDLINE=95115033; PubMed=7815459;  
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;  
RT "Characterization and nucleotide sequence of the cardiac alpha-myosin  
RT heavy chain gene from Syrian hamster."  
RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).  
RN (12)  
RP SEQUENCE OF 962-1935 FROM N.A.  
RX MEDLINE=88247788; PubMed=3380703;  
RA Jandreski M.A., Sole M.J., Liew C.-C.;  
RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin  
RT heavy chain."  
RL Nucleic Acids Res. 16:4737-4737(1988).  
CC -1- FUNCTION: MUSCLE CONTRACTION.  
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE  
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
CC SUBFRAGMENT (S2).  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: L12104; AA02313.1; -  
CC EMBL: X07273; CAA30256.1; -  
CC PIR: A28298; A28298.

DR HSP; P08799; 1MD.  
DR InterPro: IPR000048; IQ\_region.  
DR InterPro: IPR004009; Myosin\_N.  
DR InterPro: IPR002928; Myosin\_tail.  
DR InterPro: IPR001609; myosin\_head.  
DR Pfam: PF00063; myosin\_head.  
DR Pfam: PF00612; IQ\_2.  
DR Pfam: PF01576; Myosin\_tail.  
DR Pfam: PF02736; Myosin\_N.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR ProDom: PD000355; myosin\_head.  
DR SMART: SM00015; IQ\_1.  
DR SMART: SM00242; MYSC; 1.  
DR PROSITE: PS50996; IQ\_1.  
KW Myosin: Muscle protein; coiled coil; Thick filament; Actin-binding;  
KW ATP-binding; Methylation; Alkylation; Multigene family;  
KW Calmodulin-binding.  
FT DOMAIN 1 779 MYOSIN HEAD-LIKE.  
FT DOMAIN 780 809 IQ.  
FT DOMAIN 839 1934 COILED COIL (POTENTIAL).  
FT NP\_BIND 177 184 APP.  
FT DOMAIN 654 676 ACTIN-BINDING.  
FT DOMAIN 756 770 ACTIN-BINDING.  
FT MOD\_RES 128 128 METHYLATION (TR1-) (POTENTIAL).  
FT MOD\_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).  
FT MOD\_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).  
FT CONFLICT 966 966 D -> E (IN REF. 2).  
FT CONFLICT 978 978 T -> TE (IN REF. 2).  
FT CONFLICT 986 986 E -> Q (IN REF. 2).  
FT CONFLICT 1008 1014 DLQAEED -> ALEAKRT (IN REF. 2).  
FT CONFLICT 1057 1057 D -> Y (IN REF. 2).  
FT CONFLICT 1060 1060 L -> V (IN REF. 2).  
FT CONFLICT 1095 1095 D -> N (IN REF. 2).  
FT CONFLICT 1217 1217 E -> D (IN REF. 2).  
FT CONFLICT 1271 1271 D -> N (IN REF. 2).  
FT CONFLICT 1327 1327 T -> A (IN REF. 2).  
FT CONFLICT 1358 1358 C -> R (IN REF. 2).  
FT CONFLICT 1504 1504 L -> V (IN REF. 2).  
FT CONFLICT 1537 1537 M -> L (IN REF. 2).  
FT CONFLICT 1556 1556 N -> K (IN REF. 2).  
SQ SEQUENCE 1934 AA; 222928 MW; FDBAC58310B0B57D CRC64;

Query Match  
Best Local Similarity 52.2%; Score 122; DB 1; Length 1934;  
Matches 94; Conservative 64; Mismatches 139; Indels 128; Gaps 19;

QY 7 QIDELSKPL-----SDSICGVYLK--LEKSAFRLRNEF-----NVAQTALR 47  
DB 1384 ELEBAKKKLAQRLQDAEAEVAVNAKSSLEKTKRH-LQNEIEDLVADVERSNMAAALD 1442  
QY 48 KLSQN-----PSADERDALQENC-----LNKWKILSD 74  
DB 1443 KQKNFDPKILAEWKQKYEESSQSESSQKPARSLSTELFKLNAVEESLEHLEFKRENK 1502  
QY 75 SLVEQFS-----KTRDIELISWFAVQFL-LDTLLESANSLWMLADLSEKHW 123  
DB 1503 NIDQEISDLTQGLSGTSGKSHLEKIRKQLEAEKMEQLSELEBAESLF-----HEGN 1556  
QY 124 HLNVPVLPVETLKSD-----DQKGEREQA--DAKVAFPQVLGVDSSESSILYAPVLDLP 175  
DB 1557 ILRQLEFNQIKAEIERKLEKDEEMQAKRNHLRVVDSLQTSLDATRSRNEALRYKKK 1616  
QY 176 LVGEVYFFDPQ-----SARKGEISQLKSMLTYYAQ-----ERFAT---- 212  
DB 1617 MEDDLNEMETQLSHANMAAEQAQVKSLOSLLKTDQIQDDAVRANDLKENAIAYER 1676  
QY 213 ----OPKMEAKRCVQLDLSAL-----VSTKCHSLGSSQSTNFGFAK-----SL 253  
DB 1677 NNLQAELEELRAVVEQTEFSRKLADEGLIETSERVOLLSNTSLINQKKKKDADLSOL 1736  
QY 254 LTRVENALVHLSCIKLAPKRAEAKTVDEVAESSVSGELPSHDTK-HIERIPMASBOAQ 312  
DB 1737 QTEVEEA-----VQECRNAAEK-AKKAITDAMMAAEELKKEDDTSAHLERMKKNBQTI 1789



```

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEMOYOSIN (LM) AND 1 HEAVY MEMOYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (SI) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32574; AAA74199.1; -.
DR HSSP: P13538; 2MYS.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00663; myosin_head.1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR000193; MYOSINHEAVY.
DR PRODOM: PD00035; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
DR MYOSIN: Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
KW MYOSIN HEAD-LIKE.
FT DOMAIN 1 783
FT DOMAIN 1 784 813
FT DOMAIN 1 842 1938
FT NP_BIND 179 186
FT DOMAIN 658 680
FT DOMAIN 760 774
FT MOD_RES 35 774 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (MONO-) (BY SIMILARITY).
FT MOD_RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 698 698 ALKYLATION (SH-1) (BY SIMILARITY).
FT MOD_RES 708 708 ALKYLATION (SH-2) (BY SIMILARITY).
SQ SEQUENCE 1938 AA; 223064 MW; D8A8A2EC5B182626 CRC64;

Query Match
Best Local Similarity 21.3%; Pred. No. 8; 7; DB 1; Length 1938;
Matches 112; Conservative 87; Mismatches 196; Indels 131; Gaps 25;

OY 7 QIEQLSKPLSD-----DSICGYLLKLEKSAFRLRNEF-----NVAQTALR 47
DB 1388 ELEBAKKLAIORLODAEHEHVAVMACASLEKTKOR-LQNEVEDLMIDVERTAAACALD 1446
OY 48 KLSQNSADERDALQACLNKKKILSDSLYEQSKTTROIELISFVAAOFLDTTLESA 107
DB 1447 KKQRN-----FDKILAEKHKKEETHAELASQKRSRSLSTVY---FKVKNAVEES 1494
OY 108 ANSLLEWL-----ADLSEKHWDLNPVLPVETLKSDDKGR-----EODADAV 151
DB 1495 LQOLETLKRNKNLQOEISDLTQELAGSKRIHELKVKVQYQOESELOALEEYASI 1554
OY 152 KAFQVGVSESSILYAPVLOPLVGEVTFDFQSAERKEISQK-----SMLTT 203
DB 1555 -----EHEGKILR---IOLELNQVKSSEIDRKIAEKDEIDQKRNHRIVESMOQT 1603
OY 204 TVAGER-----FAIQFKME-NAKRCVTOUDRLSALVSTCHSGSSTNFGPKSLITRY 257
DB 1604 LDNEIRSDAIIIRIKKMEGLDNEIEIOLNHNRMMA---EALRNYRNTQIGLKDTQLHL 1660

```

```

OY 258 ENALVHLGKIR--LAP-----KAERK-----TVEQVAESSVSGELPSHMDTKHIER 303
DB 1661 DDLRLQGEDLKEQLAVERRANLQAEIETELRLTQGTESKRVAQEL---LD----- 1711
OY 304 IPWASQACTVSOHLHAGNLSELGNLNNMNRDLAPHLRLVSDYFPQSEPHSPISFLER 363
DB 1712 ---ASERVOL---LHTQNTSLINTFKKLETDTIS-OIOGEMEDIVQEARNAEKA---RK 1760
OY 364 AIRWGLSLPELLERMSQNGDALSTIF--NAAGLNH-LD-----QVLLP 406
DB 1761 ALTDAAMAEELKKEEDTSAHLERMKKNMEQTVKDLQHLRLDEAQLAGKKQIOKLEA 1820
OY 407 EVSTPYGIESPOTPAKPSVSDPSRSEHVSQTSVPYDQSKODK 452
DB 1821 RVNELAEVSESQ---KRNVAVKGLRKHRRVKRELTQTEEDRK 1862

RESULT 15
M10.CAEEL STANDARD: PRT: 650 AA.
ID M10.CAEEL
AC P34400;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE M19-10 protein.
GN M19-10 OR F1059.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders E., Shownkeen R.,
RA Sims M., Smallton N., Riffken A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38(1994).
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L10986; AAA28018.1; -.
DR PIR: S44806; S44806.
DR WormPep: F10E9.6; CE00150.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000159; RA_domain.
DR Pfam: PF00169; PH; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00314; RA; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
FT DOMAIN 327 437
SQ SEQUENCE 650 AA; 73318 MW; C3532BBF587D1335 CRC64;

Query Match
Best Local Similarity 19.5%; Pred. No. 2; 1;
Matches 101; Conservative 85; Mismatches 208; Indels 124; Gaps 24;

```

```

QY 7 QIBQLSKPLSDSICGVYKL---EKSAPRPLRNE-FNVAQTALRKLSONPSADERDALQ 62
Db 178 KIROALEKMEKAVTKIFVFEVEDGEALQLLIDERWTADTLKQLEKNHIA---LME 233
QY 63 EACLINRW--KILSDSLYEQSKTTROELISW-----FVAAQFLDPT 102
Db 234 DHCIVEEYFELYTKRYEDHEKVENIOM--WQODSPNKLYEMRRDPDKYAFISRPELYL 291
QY 103 TLESA-----ANSLEWLADLSEK--HMDHLNPVLPVET-----LKSDDDKGKEREQADA 149
Db 292 TPKTSDHMETPSGDQWIDYKQKFVSEYFHREPVVPPEMGGFLYLKSDGKRSWK----- 345
QY 150 KYAFAPOLYGDSEBSLILAPVULQPLVGEVT---FFDQSAERKGEISQLKSMLTTTV 205
Db 346 --KHYPVL---RPSGLIYAPKSKKPTTKDITCLMNLHNSNQVYTGIGMEKKYKSPTPWCI 399
QY 206 AQERFAPQFMENAKRCVYQOLDRLSA---LVSTKCHSLGSOSTNFGFAKSLTRVENALV 262
Db 400 STKLTALQMRRSOFIKYICAEDEMTFRKWLVALRIAKNGAE-----LLENYERA-C 449
QY 263 HLSGIRKLAPKAFAKTVQOEVAESSVSEGEPLSHMDTKHIERIPMAEQOQTVSOHLHAGN 322
Db 450 QIRRETLGPAS-----SMSAASSSTAISEVPHSL--SHQRTPSVASSIQ-LSSHM----- 497
QY 323 LSELGNLNNNRDLAFHLRLREVSDYFRQSEPHSPISFLLEKAIWMGYLSLPELLREMME 382
Db 498 -----MNPETHPLSVNVRNQ-----SPASFVNSCOO-----SHPSRTSAKLEI 536
QY 383 QNGDALSTIFNAGLNLDOVLLPEVSTPVVGIES-----PQTP-----QAKPSYSD 429
Db 537 QYDEQPTGTIKRAPLDVLRKVSRASTSSPTIQEESDSDDEFPAPPVAVSMRMPVPVT 596
QY 430 PRSVEEHVSQTSF-----VDTQSKODOKPOSSATSAL 461
Db 597 PKPCTPLTSKKAPPPPKRSDTTKLOSASAPMAKANDL 634

```

Search completed: March 2, 2003, 04:19:12  
 Job time : 39 secs

GenCore version 5.1.4.p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2003, 21:25:27 ; Search time 72 seconds  
(without alignments)  
1324.997 Million cell updates/sec

Title: US-09-915-706a-2

Perfect score: 2343

Sequence: 1 MPLSKHQIQGLSKPLSDSI.....DTQSKQDKQKQSSATSALSM 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_21:\*  
2: sp-archaea:\*  
3: sp-bacteria:\*  
4: sp-fungi:\*  
5: sp-human:\*  
6: sp-invertebrate:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*  
15: sp-virus:\*  
16: sp-bacteriophage:\*  
17: sp-archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description         |
|------------|-------|-------------|--------|-----------|---------------------|
| 1          | 252   | 10.8        | 358    | 2 Q93ED3  | Q93ED3 rhizobium 1  |
| 2          | 241.5 | 10.3        | 351    | 16 Q807V6 | Q807V6 agrobacteri  |
| 3          | 232.5 | 9.9         | 337    | 16 Q8XRS2 | Q8XRS2 ralsionia s  |
| 4          | 209   | 8.9         | 366    | 16 Q91IB9 | Q91IB9 pseudomonas  |
| 5          | 152.5 | 6.5         | 1454   | 4 Q9UPS3  | Q9UPS3 homo sapien  |
| 6          | 143.5 | 6.1         | 959    | 5 Q8TJP2  | Q8TJP2 dictyosteli  |
| 7          | 143   | 6.1         | 884    | 5 Q02239  | Q02239 caenorhabdi  |
| 8          | 141.5 | 6.0         | 344    | 16 Q91750 | Q91750 pseudomonas  |
| 9          | 141.5 | 6.0         | 1501   | 5 Q22699  | Q22699 caenorhabdi  |
| 10         | 140.5 | 6.0         | 3742   | 8 Q9TKW8  | Q9TKW8 nephroselimi |
| 11         | 136.5 | 5.8         | 7201   | 5 Q9VMT8  | Q9VMT8 drosophila   |
| 12         | 134   | 5.7         | 1051   | 3 Q60039  | Q60039 emericella   |
| 13         | 129.5 | 5.5         | 630    | 5 Q25613  | Q25613 onchocerca   |
| 14         | 129.5 | 5.5         | 1133   | 5 Q21022  | Q21022 caenorhabdi  |
| 15         | 128   | 5.5         | 1978   | 4 Q15154  | Q15154 homo sapien  |
| 16         | 127.5 | 5.4         | 2471   | 12 Q9YTK3 | Q9YTK3 ateline her  |

|    |       |     |      |           |                    |
|----|-------|-----|------|-----------|--------------------|
| 17 | 127   | 5.4 | 1935 | 6 Q9BE39  | Q9BE39 bos taurus  |
| 18 | 126.5 | 5.4 | 752  | 4 Q13597  | Q13597 homo sapien |
| 19 | 126.5 | 5.4 | 1023 | 16 Q92B64 | Q92B64 listeria in |
| 20 | 126   | 5.4 | 613  | 2 Q9XD53  | Q9XD53 moraxella c |
| 21 | 125.5 | 5.4 | 1992 | 13 Q04834 | Q04834 xenopus lae |
| 22 | 125   | 5.3 | 1194 | 3 Q42649  | Q42649 schizosacch |
| 23 | 125   | 5.3 | 1935 | 6 Q9GKR1  | Q9GKR1 sus scrofa  |
| 24 | 124.5 | 5.3 | 550  | 3 Q59725  | Q59725 schizosacch |
| 25 | 124.5 | 5.3 | 1964 | 13 Q93522 | Q93522 xenopus lae |
| 26 | 123.5 | 5.3 | 1051 | 2 Q49524  | Q49524 mycoplasma  |
| 27 | 123   | 5.2 | 611  | 4 Q14905  | Q14905 homo sapien |
| 28 | 123   | 5.2 | 800  | 10 Q94GM1 | Q94GM1 oryza sativ |
| 29 | 123   | 5.2 | 1935 | 4 Q9H1D5  | Q9H1D5 homo sapien |
| 30 | 123   | 5.2 | 2471 | 13 Q9PDN8 | Q9PDN8 xenopus lae |
| 31 | 122   | 5.2 | 1344 | 2 Q49545  | Q49545 mycoplasma  |
| 32 | 122   | 5.2 | 1370 | 5 Q9VD13  | Q9VD13 drosophila  |
| 33 | 122   | 5.2 | 2735 | 5 Q76416  | Q76416 caenorhabdi |
| 34 | 121.5 | 5.2 | 1179 | 16 Q97OG7 | Q97OG7 streptococc |
| 35 | 121.5 | 5.2 | 1871 | 5 Q9NCL3  | Q9NCL3 drosophila  |
| 36 | 121.5 | 5.2 | 3259 | 4 Q14789  | Q14789 homo sapien |
| 37 | 121   | 5.2 | 765  | 11 Q9D2N7 | Q9D2N7 mus musculu |
| 38 | 121   | 5.2 | 1058 | 10 Q9FLR5 | Q9FLR5 arabidopsis |
| 39 | 121   | 5.2 | 2501 | 5 Q9NCW7  | Q9NCW7 rattus norv |
| 40 | 120.5 | 5.1 | 862  | 11 P70609 | P70609 drosophila  |
| 41 | 120.5 | 5.1 | 1048 | 5 Q9W018  | Q9W018 drosophila  |
| 42 | 120.5 | 5.1 | 1294 | 5 Q9VXU1  | Q9VXU1 drosophila  |
| 43 | 120.5 | 5.1 | 1398 | 5 Q96D00  | Q96D00 drosophila  |
| 44 | 120.5 | 5.1 | 1939 | 6 Q9TV61  | Q9TV61 sus scrofa  |
| 45 | 120   | 5.1 | 391  | 16 Q98IM6 | Q98IM6 rhizobium 1 |

## ALIGNMENTS

RESULT 1  
ID Q93ED3 PRELIMINARY; PRT; 358 AA.  
AC Q93ED3:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE IMPA.  
GN IMPA.  
OS Rhizobium leguminosarum (biovar trifolii).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID:386;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bladergroen M.R., Badelt K., Stronk O.P., Lugtenberg E.J.J.,  
RA Spalink H.P.;  
RT "An avirulence locus of a symbiotic Rhizobium leguminosarum strain is  
RT involved in temperature-dependent protein secretion.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF61470; AAL17808.1; -  
SQ SEQUENCE 358 AA; 39039 MW; 9F5C107C58D22395 CRC64;

Query Match 10.8%; Score 252; DB 2; Length 358;  
Best Local Similarity 23.9%; Pred. No. 4.1e-09;  
Matches 96; Conservative 71; Mismatches 170; Indels 64; Gaps 14;  
QY 8 IEQLSKPLSDSICGVVLEKLSAFRLRNENVAQTALRKLQNPASDEKDALQDEACLN 67  
3 VREITIDPLEQNHPCGENVR-SNTAFREIYRIKQARNNAKFAEKSIIPGETITTLAPA-- 58  
QY 68 KWKILSDSLVEQFSKTRTDIELISFVAAGFLDPTTLESANSLSEWLAD-----LSEK 120  
59 -WHQVSNIGQLLSKSKDIEVLAWLAQAQRL-----RGPSGLHDVYVATVSLDK 109  
QY 121 HWDHLNPLYLVETLTKSDDKKEREQADAKYKFAFPOLVGSEESIIYAPYQLPLVG-- 178  
110 HFDALHSI-----GDGD-----VEERFAPAGLNGVGEGTLIAQIRLTLIPGCK 155

```
QY 179 --EYTFDFQSAERKKGELISQKSLMTTVAQERRAIOFKMNAKRCYTQDLRLSLAVSTK 236
DB 156 FAQSFSLWDFQISQRPNEFERRKLQQAAGVAMSAVLDVLTGCLAAFPBWEIIDEH 215
QY 237 CHSLGOSTNFGFAKSLTLTRENALVHLSGIKLAPKAEAKTVEQVAVSSVSEGLPESHM 296
DB 216 CGDQAPFSSN---TRNVLOEASAIRMLAGIE-AP-TEAPLAQALASQSE-PLPA-S 268
QY 297 DTKHIERIPMAEQAVTSQHLHAGNLSLGNLNMNMDLAFHLLREVSDFRQSEPHSP 356
DB 269 ETDSESAITGRVSAEMT-----RSREAFPELLIAVARFRFRTEPHSP 311
QY 357 ISFLLEKAIKRWGYSLEPELLREMSQNGDALSTIFNAGL 397
DB 312 ISMSIETLVRRGRMDFELLAEELPEQ--QTRNAVLTAAGI 350

RESULT 2
ID 0807V6 PRELIMINARY; PRT; 351 AA.
AC 0807V6;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein Atu4343.
GN Atu4343 OR AGP_L_1042.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_Taxid=176299;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; Pubmed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitaajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Kester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; Pubmed=11743194;
RA Goodner B., Hinkle G., Gatlung S., Miller N., Blanchard M.,
RA Ourrelio B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlet K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE009363; AAL45137.1; -.
DR EMBL: AE008251; AAK69097.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 351 AA; 38490 MW; 85460F21E1C63FC7 CRC64;

Query Match 10.3%; Score 241.5; DB 16; Length 351;
Best Local Similarity 23.6%; Pred. No. 2e-08;
Matches 99; Conservative 64; Mismatches 162; Indels 85; Gaps 15;
```

```
DB 49 PQDNLKISSSDSVNGLQIITYSEKDVLELAMLAEVSLRL-RGPHGLREITYELCGDLF 107
QY 119 EKMDHLNPLYVETLKSDDDKGERQADAKYAFQLVGDSSESSITLAPVLQPLV- 177
DB 108 YNHMDSLRSI-----SDND-----EKKFAPFAGLNGIGSEGT-LVQPLRLASLIP 152
QY 178 ----GEYTFDFQSAERKKGELISQKSLMTTVAQERRAIOFKMNAKRCYTQDLRLSLAV 233
DB 153 GKGKGEHSLWDFQISQRPNEFERRKLQQAAGVAMSSHLAAVNTCLSSFPAITAVL 212
QY 234 STKCHSGOSTNFGFAKSLTLTRENALVHLSGIKLAPKAEAKTVEQVAVSSVSEGLPESHM 290
DB 213 SERGQAPFSSN-----IRNLTLEAA-----AAIRTLGRDQDPAP----- 249
QY 291 ELPSHMDTKHIERIPMAEQAVTSQHLHAGNLSLGNLNMNMDLAFHLLREVSDFRQ 350
DB 250 -----VEQTPAIA--AGTDESGQSAARTSPASPEGISSRDEAFETLLSVAFRFR 297
QY 351 SEPHSPISFLLEKAIKRWGYSLEPELLREMSQNGDALSTIFNAGL 397
DB 298 TEPHSPISLTIETLVRRGRMDFELLAEELPET--QARNNAVLTAAGI 342

RESULT 3
ID 08XRS2 PRELIMINARY; PRT; 337 AA.
AC 08XRS2;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical protein RSP0759.
GN RSP0759 OR RS01949.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_Taxid=305;
RN
RP SEQUENCE FROM N.A.
RX STRAIN=GM11000;
RX MEDLINE=21618187; Pubmed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brothier P., Camus J.C., Catorlico L.,
RA Chandler M., Choisine N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL: AL646080; CAD17910.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 337 AA; 37094 MW; C1EE68258B839897 CRC64;

Query Match 9.9%; Score 232.5; DB 16; Length 337;
Best Local Similarity 24.9%; Pred. No. 7.7e-08;
Matches 99; Conservative 60; Mismatches 127; Indels 111; Gaps 18;
```



```

QY 320 -----AGNLSELGMLNNNNRDLAFHLREVDYFROSE 352
      : || E L : : |
Db 675 NPKRESVITGHFEKPEKPKIKIDSEPLCLSDNLMLHQDFDPLN -----VQELSENF----- 723
QY 353 PHSPISTLLEKAIKRWGLYSLEPELLREMSB-ONGDALSTFFNAAGLNHLD----- 401
      ||| : | | : : : | : |
Db 724 -----LELQEKNLKLGSLSKSEKHINDLQTELKNAGTEAMLETSCNSNDLDELQAEKNP 778
QY 402 -----QVLL-----PEVSTPIVGIESTPOTPAKRSVDPSRY 433
      : | : | | | : : | | |
Db 779 GMSLLQENSTKGDGDDIVMLTGTLTSLSSQSEVQVPPTSFETETPRRVPDLSLPIQG 838
QY 434 EEHVSQTSPYDTOSKODKPOSSATSALS 462
      | : : | : | : | :
Db 839 E---TOPPTCLDVIVPEDCLNHQDISPAVY 864

RESULT 6
Q871P2
ID 08T1P2 PRELIMINARY; PRT; 959 AA.
AC 08T1P2;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE TATA element modulatory factor.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
  [1]
RN SEQUENCE FROM N.A.
RP STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Liehmann R., Baumgart C., Parra G., April J.F., Guilo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Neigel A.A.,
RT Sequence and Analysis of Chromosome 2 of Dictyostelium.
RL Submitted (MAR-2002) to the EMBL/Genbank/DBD databases.
SQ EMBL, ACL16330; AAL96709.1;
DO SEQUENCE 959 AA; 110909 MW; 151551A907D72335 CRC64;

```

[illegible]

Db 889 KEGEAMSLQAOVQSLNHSRKKLEDEL 914

RESULT 7

ID 002239 PRELIMINARY; PRF: 884 AA.

AC 002239: 002269;

DT 01-JUL-1997 (TREMBLrel. 04, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE E01G4.6 protein.

DN E01G4.6.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_Taxid=6239;

RN [1]

RA SEQUENCE FROM N.A.

RA Sins M.;

RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RA SEQUENCE FROM N.A.

RP MEDLINE=94150718; PubMed=7906398;

RX Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A., Gardner M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gaxton A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Keshaw J., Kirsten J., Laister N., Latteille P., Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R., Smalton N., Smith A., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sprat J., Wooldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";

RT Nature 368:32-38(1994).

RL Nature 368:32-38(1994).

RN [3]

RA SEQUENCE FROM N.A.

RP Sins M.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: 283223; CAB05719.1; -.

DR EMBL: 281538; CAB05719.1; JOINED.

DR EMBL: 281538; CAB04387.1; -.

DR EMBL: 283223; CAB04387.1; JOINED.

DR InterPro: IPR001211; PhospholipaseA2.

DR PROSITE: PS00118; PA2\_HIS; UNKNOWN\_1.

SO SEQUENCE 884 AA; 99603 MW; 72A1610965C9B30D CRC64;

Query Match 6.1%; Score 143; DB 5; Length 884;

Best Local Similarity 20.7%; Pred. No. 0.3;

Matches 103; Conservative 99; Mismatches 165; Indels 130; Gaps 25.

QY 27 LEKSAFRLRNDENFVAOALRKLKSLNSPADERDALQECILNWK-----RIKLSLVEQESK 82

Db 218 LRPDAVITARITRLRLATLSTATLSR-----RERFLKYSIROIYVAQNLTGQOOR 268

QY 83 TTRDIELISWEVAAOFLDITLLESANSLSEWLADSLSEKHWMDLNPVLPYEYLSDDDKGK 142

Db 269 LTRPEYEDLVTESSKSLAVEAPEETTGQDETTDASE-----VTTTKAVEARE 317

QY 143 E-RECAADAKVAFAFQVODSESSSLIVAPVLO-----LPLVGEVTFPD-FQ 186

Db 318 EYTEATEATEA---PVATTKESSSEMHVNTIRNMLRSASEKDSLKYVTLISGKSESLFE 374

QY 187 SAERK-----GEISOLKSMLETTTYAO-----ERFAIOPKMEAKKRCYQL 226

Db 375 LAEOQKLTLLTSKFEDEKLSSKAKMLDLINLEALSEKSEKGEIDQAMEKFEKPKSELVAME 434

QY 227 DRLSALVSTKCHSLCSQSTNFGFAKSLITR-----VENALVHLSGIPLKPAEAKTVQED 281

Db 435 DKDPAVAVETIDSLKHKAEKLAFTIYSRVNVEENM---IEKEVPEKKAEEKVKVEED 491

QY 282 ---VAESSVSE---GELPSHMDTKHIERIPMASEQAQTYSOHLHAGNLSLSEGLNLMN--N 333

```
DB 492 VKAAVEEKKEKKPKLP--MKIEKLEK-FVDTK-----SENHBLKVLDD 534
QY 334 RDAFHLRLREVSDFYFROSEPHSPISFLLEK---AIRMGYLSLPELLREMSSEQ---NGDA 387
DB 535 KERALLVESIEIKTAETKPKVE-SFRSEETVVAID---DMPALKEEESAERKETTGE 589
QY 388 LSTIFNAAGLNHLDOVLLPEVSTPTVIGIESPOTPAKPSVSDPRSYEENVSQTSPTDQT- 446
DB 590 TTT-----EAAVETTEASTPRKPAKPELLS--NLEDVLTITTP-ETET 630
QY 447 ---SKODKPOSSATSA 460
DB 631 IEGSGEREETPTTAPAA 647

RESULT 8
QY 091750 PRELIMINARY; PRT; 344 AA.
ID 091750;
AC 091750;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE Hypothetical protein PA0082.
GN PA0082.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OC NCBI_TaxID=287;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stoyer C.K., Pham X.-Q.T., Ervin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Gardner R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RA opportunistic pathogen.";
RL Nature 406:959-964 (2000).
DR EMBL; AEO04447; MAG03472.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 344 AA; 36679 MW; 3321BAC4C8DB025D CRC64;

Query Match 6.0%; Score 141.5; DB 16; Length 344;
Best Local Similarity 21.1%; Pred. No. 0.1;
Matches 89; Conservative 70; Mismatches 144; Indels 119; Gaps 18;

QY 15 LSDDSICGVYLYLEKSAFRLRNEFNVAQTALRKLKSG-NPSADERALQACLNKKILS 73
DB 11 VSPDSFGDDL-----EYDAFLELERIAGOGPEROMGADVLAERPEMVR 58
QY 74 DSLYEQSKTTREDIELISWFAQFLDITLESANSLEWLADISEKHMDLNPVLPVET 133
DB 59 ALASELFGK-SKDLRVANLLLOSVALD-GLDLADGULLVRELIGQYMGVPL----- 111
QY 134 LKSDDDKGEREQADAKVKAFFOLVGDSESSITLYAVLQL---PLVGEVTF--FDEQS 187
DB 112 LDADD-----NDPTFRINALTGLVAE-----PLLDLWVAIPLVRSRAGPVLLRA 157
QY 188 AEAKGEISOLSKMLTTTVAOERFAIOFKMENAKRCYTQDLRLSALVSTKHSIGSOSTNF 247
DB 158 ALNAAAGLQRPAS---ETLSPEQTAGAFADADA-----DALATRALDGADEHAL 204
QY 248 GFAKSLTRVENA-----LVHL-----SGIKLAPKAEAKTYEQEVAES 285
DB 205 AITSSGAERYGSAOGLDGLRLQLRALQVFPFLYGRQGESLAGAEN-VADEQVGAA 263
QY 286 SVSEGLPSIMDTKHIERIPMASEQOAVYSQHLHAGNLSELGNLNNKRDLAFFHLLREVS 345
DB 264 PVAAVAAPA-----PRAS-----GEIA-----NREDVLRQDLRL 293
```

```
QY 346 DYFROSEPHSPISFLEKAIKRMGYLSLPELLREMSFONGDALSTIFNAGLNHLDOVLL 405
DB 294 EYVVRHPPSPVYVLLKRAKTLVTADPAEIVRLIPD-----GISOFETLRG 340
QY 406 PE 407
DB 341 PE 342

RESULT 9
QY 022699 PRELIMINARY; PRT; 1501 AA.
ID 022699;
AC 022699;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 173.3 kDa protein.
GN T23F2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None.
RA "Genome sequence of the nematode C. elegans a platform for
RA investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN 12)
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du Z.;
RA "The sequence of C. elegans cosmid T23F2.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN 13)
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RA "Direct Submission.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39649; AAB0384.2; -.
KW Hypothetical protein.
SQ SEQUENCE 1501 AA; 173309 MW; 271052697CE0CF97 CRC64;

Query Match 6.0%; Score 141.5; DB 5; Length 1501;
Best Local Similarity 20.0%; Pred. No. 0.79;
Matches 112; Conservative 89; Mismatches 168; Indels 191; Gaps 26;

QY 9 EQLSKPLSDSICGVYLYLEKSAFRLRNEFNVAQTALRKLSONPSADERALQACLNK 68
DB 619 QQLSEHESD-----FLSTQOT-LHDLKKEYSAKNNTL-----VDKFEVEEILLAK 663
QY 69 WKLLSDSLYQFSK---TTRDIELISWFAQFLDITLESANSLEWLADISEKHMDH 124
DB 664 TELV-DALTKQLENIRKQQRLELS-----LQSERDQYKKSLSEMTFTAEC--- 708
QY 125 LNVLPVETLKSDDDKGEREQADAKV---AFQO-----LVGDS----- 161
DB 709 -VPILEAEILQLSKDK---NETIARKKHQDEYFEDBELAKLLMDSNNIKKERRDYLEHI 763
QY 162 --ESSITLYAVLQLPLVGEVTFPFQOSAEK-KGEISOLSKMLTTTVAOERFAIOFKMEN 218
DB 764 RANESMI-----ERLKLLEISGLKDLLENOKMQAHLO---KSEL 798
QY 219 AKRCVQOLDRLSALVS-----TKHSIGSOSTNGFAK----- 251
DB 799 EKKLLSIDHVSQLOSRVNHSDQDVCEQALPIRINKYGGKPNVKKETIIEGALPDEN 858
QY 252 -----SLITREVALVHL-----SGIKLAPKAEAKTYEQEVAES 286
```

Db 859 EERLRICKAELETTTRQVTVLQOKLVSIIQOQSSQIKRKRIAVEDSNKNTVHTEDESK 918  
QY 287 VSEGELEPSHMDTKHIERIPASPEOQTVSOHLHAGNLSELCN-----LNMMNR 334  
Db 919 MKREVEL---KNTLEMERID-SLEAEFVASSIEKSRIOKLVNEDNLKOKLDNDMSYSK 974  
QY 335 DLAF-----HLREVSDFYRQSEPHSPISFLEKAIKRWGLSLE-LLREMMSEON-GD 386  
Db 975 EKQWLMQIRINLEKDNSELQKOIOPSS-----EKSL-----SLNKGLTKRTNSEPDGCD 1024  
QY 387 ALSTIFNAAGLNHLHDQVLLREVSPTVYGIESPQ-----TQAKPSVSDPR 431  
Db 1025 DMSIEGAGSTNESADPMFVSSTAIPVLSLKSPQFSQLADVLNLVRSDELQVLTLEIPE 1084  
QY 432 SVEEHVSQTSPPVDTSQKOD 451  
Db 1085 AAKQEPQMSLEKSTKNQV 1104

## RESULT 10

Q9TKW8 PRELIMINARY; PRT; 3742 AA.  
ID O9TKW8  
AC O9TKW8  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Hypoetical chloroplast RF2.  
GN YCF2.  
OS Nephroselmis olivacea.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;  
OC Chlorodendriales; Chlorodendraceae; Nephroselmis.  
ON NCBI\_TaxID=31312;  
RX MEDLINE=99398694; PubMed=10468594;  
RA Turmel M., Otis C., Lemieux C.;  
RT "The complete chloroplast DNA sequence of the green alga Nephroselmis  
olivacea: insights into the architecture of ancestral chloroplast  
genomes.";  
RT Proc. Natl. Acad. Sci. U.S.A. 96:10248-10253(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Turmel M., Otis C., Lemieux C.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF137379; AAD54848.1;  
DR InterPro: IPR003593; AAA\_ATPase.  
DR InterPro: IPR003959; AAA\_ATPase-centr.  
DR Pfam: PF00004; AAA\_1.  
DR SMART: SM00382; AAA\_1.  
KW ATP-binding; Chloroplast.  
SQ SEQUENCE 3742 AA; 436427 MW; 1F45143EDBC89275 CRC64;

Query Match 6.0%; Score 140.5; DB 8; Length 3742;

Best Local Similarity 20.7%; Pctd. No. 3.3; Indels 167; Gaps 23;

Matches 101; Conservative 65; Mismatches 154; Indels 167; Gaps 23;  
QY 9 EQLSKPLSDSICG---VYLKLEKSAPRL-----RNEFNVAQRTLRKLSQNSADERDAL 61  
Db 283 EELQKQWEYVSYLGNSTYEL-ITGRFRYPGKQKQREQQAIDAKDLAKKQKEEQSK 341  
QY 62 QEACLNK-----WKILSDSYEOPSKTTRIDIELISWFAAQFLIDTTLESANSL- 112  
Db 342 KEKLARLIDGQNM-----FDKKQTKYELKALEMAEYWKEDQEAQAKDLEHKLN 391  
QY 113 ---WLAELS-----EKHM-----DHLNRPVL-----VETLKSDDK 140  
Db 392 RQAFATLTMDKVAEDRYWEESYLQYPVAGHKDLSLSPDLPSMTNRNDELNSVSEA 451  
QY 141 GKREKQADAKYKAFQVGVGSESSILYAPVLQPLVGEVTFPFOQAEKRGKETSOLKSM 200  
Db 452 LKNRRRDAAHPELHRTIGTERLESV--QVWL-----EFT--EFAINER-----IRTL 496

QY 201 LTTVAQERPAIOF--KMENAKRCVTOQLDRLSALVSTKHSLSQSTNGFASLLTRVE 258  
Db 497 YTTSLSSSHADQFLVQETFLQHLROLDEV-----AKELVDTR 536  
QY 259 NALVHLSGITLAKRAEKTYEQEVAESVSEGELEPSHMDTKHIERIPASPEOQTVSOHL 318  
Db 537 GSULD--DGIETLEPLLSRREE-----VSRV-MSLEDYERLQNLH 573  
QY 319 HAGNLSELGNLNMNNDLAFHLREVSDFYRQSEPHSPISFLEKAIKRWGLSLELLRE 378  
Db 574 -----RQDSTSP-----HLSKLTSTA 591  
QY 379 MMSQNDALSTJFMAAGLNHLHDQVLLREVSPTVYGIESPQTPQAKPSVSDPSVEBVS 438  
Db 592 VLSNPSPLMQGI---PATNPMLLVYRGDIKSPRVGLRAGLTG--RPSREAPDSSHNWI 646  
QY 439 QTS-PVD 444  
Db 647 ITSLPVD 653

## RESULT 11

Q9VMT8 PRELIMINARY; PRT; 7201 AA.  
ID Q9VMT8  
AC Q9VMT8  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE MSP-300 protein.  
GN MSP-300 OR CG18251 OR CG18252.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Pphyroidea; Drosophilidae; Drosophila.  
ON NCBI\_TaxID=7227;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer J.R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Abmayri A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Burlis R.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.L., Nelson K.A., Nixon K., Nuskern D.R., Pacleby J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheer F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spires E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swierkas R., Tecior C., Turner R., Venter R., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan K., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003608; AAF52223.2; -.  
 DR HSP: P46939; 1BDH.  
 DR Flybase: FBgn0010070; Msp-300.  
 DR InterPro: IPR001715; Calponin-like.  
 DR InterPro: IPR000345; Cytc\_heme\_bind.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PF00307; CH; 2.  
 DR Pfam: PF00435; Spectrin; 39.  
 DR SMART: SM00033; CH; 2.  
 DR SMART: SM00150; SPEC; 36.  
 DR PROSITE: PS00021; CH; 2.  
 DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 SQ SEQUENCE 7201 AA; 826597 MW; 3084D2637DE94F2 CRC64;

Query Match 5.8%; Score 136.5; DB 5; Length 7201;  
 Best Local Similarity 22.0%; Pred. No. 15;  
 Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;

QY 30 SAFRPLRNF-FNVAQTA---LKKLS-----QNP SAD---ERD---ALQEA---CLNKK 70  
 DB 1263 SEFVDQKNDFSMQLQTAETELRSLTPLOTPKNSQDLKSRDLNVQLQASHQLPLH 1322  
 QY 71 ILSDSLYEQFSKTRTDIELISWFAAQFLDITLESANSLLEMLADLSKNDH----- 124  
 DB 1323 ALKSELAPLAPDCKRPT-LEKEVTEYEMKFENMEHVKNRGVGLSEYSK-NNNYTRRLA 1380  
 QY 125 -----LNPLP---VETLSDDDKEREQADAKVKAFFOLVGD-SESSILYAPVLQ-LP 175  
 DB 1381 ELQEMANKVAPKNIEALQSEDLTPREVR---VKVQAFKRILLDRMKQDILLAADASLAP 1437  
 QY 176 LVGEVTFEFQSAERKGEISQLKSMLT----- 203  
 DB 1438 KEGNIA---EAKRLKEGTEIKLQEVLSAINRVNDHQQAQVQEDLVNMQQFOAGLIQIKPA 1493  
 QY 204 -----TYAERFALQFMENNAKRCVTOIDRLSALVSTKCHSLGSO 243  
 DB 1494 VQOSEKVVNNVSKPTSLSEAYVAMQAOQFEQ---COBOLDKLHG-TSNISHMLCK 1548  
 QY 244 STNFGFASLITRVEN--ALVHLSGIKLPKAEA---KTVEQEA--ESSVSEGE--- 291  
 DB 1549 -TN--APDELAMSRMVAHENAQAASAKLEKLVANMKSPDADAKLEDVWGQGEQOM 1604  
 QY 292 --LPSHMDTKHIERI-----PMASEDAQTV-----SQHLHAGNLSL- 326  
 DB 1605 SRPVALNTPHIDKLEKELVKLSFNNESIOOAKLVTLGQVADQISLHAFEGGAALKD 1664  
 QY 327 -----GNLNNNRDLAFHLREVSDF--ROSEPHSPISFLLEKAIIRNGYLSLPELLRE 378  
 DB 1665 RVNOMKAGKLOKLSLEATRGH-INEVSDALITSRODFNAKLNF---SNM---MGOLRN 1713  
 QY 379 MMS-----EQNGDALST---IFNAGLNLHLDQ-----VLREVPSTPVIGISPPOTPOAKPSV 427  
 DB 1714 QTVQVEINPEVEVETSLVHTHALLDEHADKPSFNALYDEVQALAG---ATPESNML 1769  
 QY 428 SDP-----RSVEHVSQTSV-----DTOS-----KQOKPOSSATSAL 461  
 DB 1770 NDAYTALVNVYONLETNMLQKAALEKWTLLGKMKNDTESHNLVTKHQLDKREGPAEEL 1829  
 QY 462 S 462  
 DB 1830 S 1830

RESULT 12  
 060039 PRELIMINARY: PRT: 1051 AA.  
 AC 060039:  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE Anucleate primary sterigmata protein APSB.  
 GN APSB.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxID=5072;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=99140441; PubMed=10094631.  
 RA Suelmann R., Slevers N., Galetzka D., Robertson L., Timberlake W.E.,  
 RA Fischer R.;  
 RT "Increased nuclear traffic chaos in hyphae of *Aspergillus nidulans*:  
 RT molecular characterization of apsb and in vivo observation of nuclear  
 RT behaviour.";  
 RL Mol. Microbiol. 30:831-842(1998).  
 CC -I- FUNCTION: INVOLVED IN REGULATION OF NUCLEAR MIGRATION. MAY BE  
 CC -I- INVOLVED IN REGULATING NUCLEAR POSITIONING.  
 CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -I- SIMILARITY: SOME TO TRICHOHYALIN.  
 DR EMBL: AJ003163; CAA05918.1; -.  
 KW Coiled coil.  
 FT DOMAIN 10 200 COILED COIL (POTENTIAL).  
 FT DOMAIN 239 285 COILED COIL (POTENTIAL).  
 FT DOMAIN 325 743 COILED COIL (POTENTIAL).  
 FT DOMAIN 787 824 COILED COIL (POTENTIAL).  
 FT DOMAIN 857 878 COILED COIL (POTENTIAL).  
 FT DOMAIN 950 1004 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1051 AA; 120996 MW; 926201A9E208847A CRC64;

Query Match 5.7%; Score 134; DB 3; Length 1051;  
 Best Local Similarity 19.2%; Pred. No. 1.5;  
 Matches 86; Conservative 62; Mismatches 134; Indels 166; Gaps 17;

QY 7 QIEQLSKPLSDSICGVYKLEKSAERPLRNEFNVAQTAIRKLSQNP SADERDALQ---- 62  
 DB 416 KIEBCOR-LTED-----LRTODNLRALQAEKRSASEGIIRLE-----DAQNNLQRYKA 464  
 QY 63 -----EACLNKKKILSDSLYEQFSKTRTDIELISWFAAQFLDITLESANSLLEMLADL 117  
 DB 465 VQGLDQCNEMESLEKSLYEANSKYOR-----LTVQIESSQNEIAFLRE- 509  
 QY 118 SEKHMDLNPVLVETLSDDDKEREQADAKVKAFFOLVGDSESSILYAPVLQ-PLV 177  
 DB 510 -----SEKDKTELEGRL-----EDDGKIK-----IGLSEELKTYRSLQ----- 533  
 QY 178 GEVTFEFQSAERKGEISQLKSMLTITVAERFALQF---KMNNAKRCVTOIDRLSALV 233  
 DB 534 -----SEKDKTELEGRL-----AEERYORBVGSKERQEVORIMNELNREVSA 578  
 QY 234 STKCH---SLGSQSTNFGFASLITRVENALVHLSG-----IKLAPAEAK 276  
 DB 579 KECKRKLKNLSAQELIETWKEKRLDLENMNRKLTGLDGRSSLIANIMKLOKLEST 638  
 QY 277 TVEQEVAESSVSEGE-----LPSH-MDTKHIERIPMASEQA-----QTVSOH 317  
 DB 639 ALELESTRSLDEKETLLNRRALLSHGLESKLSLELLEREQAARAKQSTEQALKSH 698  
 QY 318 LHA-----GNLSLGNLNNNRDLAFHLREVSDFRQSEPHSPISFLLEKAIIRMGV 369  
 DB 699 HQASRTITONNSRIELENAARNODRK-----RF 726  
 QY 370 LSLPELLREMSRONGDALSTIFNAGL 397  
 DB 727 TSLDQDFEQLNERNSMMLTWKRLSGM 754

RESULT 13  
 025613 PRELIMINARY: PRT: 630 AA.  
 AC 025613:  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE Myosin-like antigen (Fragment).  
OS Onchocerca volvulus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea.  
OC Onchocercidae; Onchocerca.  
OX NCBI\_TaxID=6282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90301142; PubMed=2194123;  
RA Ngozi E., Erondu N.E., Donelson J.E.;  
RT "Characterization of a myosin-like antigen from Onchocerca volvulus.";  
RL Mol. Biochem. Parasitol. 40:213-224(1989).  
DR EMBL: M30398; AAA29413.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 630 AA; 73919 MW; CF093116A40D889F CRC64;  
  
Query Match 5.5%; Score 129.5; DB 5; Length 630;  
Best Local Similarity 22.7%; Pred. No. 1.5; Indels 81; Gaps 15;  
Matches 84; Conservative 64; Mismatches 141;  
  
QY 35 LRNEFVAQTALRKLSNPSSADERDALQEAQLNKMKILSDSLYQFSKTRDIELISMFV 94  
DB 239 LRSLKLNNAERMADL-QN-----RDSILRENNNDMKESDALNNEIDLR-LRD-ELLSVRR 290  
QY 95 AAGFL--DTTLESAANSELEMLADISEKHMDLNPVLPVETLKSDDKGKEREQADAKV 151  
DB 291 DAKEIKRNYNTDLOTARNEIKILSTNNEMKSQLN-----AARDKI 331  
QY 152 KAFQVLGDSSESSILVAPVLQPLVGEVTFPFQSAERKGEISQLKSMLTITVAQERFA 211  
DB 332 NSLAKVTIDQONKI-----RDLTGEVHHLEBELKDAKGNVANLSELDLT---RER 379  
QY 212 IOFEMENAKRCVYQDLRLSALVSTKCHSLGSSQSTNFGFAKSLTRVENALVHLSGRLAP 271  
DB 380 IHLLEGONASLQTELNKIK-----GDIDSLFG-----ENDM-LTKAESN 418  
QY 272 KAEKTYEQEVAESSVEGELPSHMDTKHIERIPMASEOQVTSQHLHAGLSLGLUNN 331  
DB 419 EAELDRLKOKLQRSIENAKKYSDALDKLREPYRL-----QNLVREKIKQAEMLTQ 469  
QY 332 MNRDLAFHL-----LREVSDFRQSE-PSHSIFLLEKAIKRMGYLSLPELLREPMSEON 384  
DB 470 AVQDLSESLNDSRELDAIDDKLIASGDRNALNSEYKLOHEVQPMREQLRK--TDYE 527  
QY 385 GDALSTIFNA 394  
DB 528 QAALSDLVNA 537  
  
RESULT 14  
Q21022  
ID 021022 PRELIMINARY; PRT; 1133 AA.  
AC 021022: 021071;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-MAY-1999 (TREMblrel. 21, last sequence update)  
DT 01-JUN-2002 (TREMblrel. 11, last annotation update)  
DE F59A2.6 protein.  
GN F59A2.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RL lightning J;  
Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Z34801; CAAB4332.1; -.  
DR EMBL: Z66514; CAAB4332.1; JOINED.  
DR EMBL: Z66514; CAAG1344.1; -.  
DR EMBL: Z34801; CAAG1344.1; JOINED.  
DR InterPro: IPR000237; GRIP\_domain.  
DR Pfam: PF01465; GRIP; 1.  
SQ SEQUENCE 1133 AA; 129239 MW; 92015E790A9DE01F CRC64;

Query Match 5.5%; Score 129.5; DB 5; Length 1133;  
Best Local Similarity 21.3%; Pred. No. 3.4;  
Matches 110; Conservative 77; Mismatches 184; Indels 145; Gaps 23;  
  
QY 9 EQLSKPLSDSICGVYIKLEKSAFRLPRNFNVAQTLAKRLSONPSADE---RDALQEA 64  
DB 116 EQLK-----AMEKLSNPONTILDEVTKLEQ--SEEEVLAARGAIOE- 155  
QY 65 CLKMKWILSDSLYQFSKTR---RDIELISFVAQAQFLDPTLESANSLLEMLADISEKH 121  
DB 156 -----LTKLESEKETSTAKTELEAVS-----KKLDSSTSLKESDME-- 196  
QY 122 WDHLNPVLPVETLKSDDKGKEREQADAKVKAFFQVLGDSSESSILVAPVLQPLVGEVTF 181  
DB 197 -----AMKIQLINCEKQKQAEVALLKQKLEVEKKNMSDVE-----VQKOLLEST 241  
QY 182 FFDF-----QSARKGEISQK-----SMLTTVAQERFAIQ---FKMENA 219  
DB 242 TSEKQHAEEAEIVYKQLEEAQSIENLKDAENERNLKTALSDSSAISLETKQMEAA 301  
QY 220 KRCV-----TQDLRLSALVSTKCHSLGSO-----STNFGFAKSLTRVENAL 261  
DB 302 KKELEASERKSELREQMDKIQ-----KVHNAQEDIQKIQKTELEMAKIAKSTEDKL 356  
QY 262 V-HLSGIKLAPRAEAKTVEQEV-----AESVSEGEPLSHMDTKHIERIPMASEOQV 313  
DB 357 ARQLAGELNNAKEDLKVVEEKHTGIRQAQALDPADEKVVYKQLQERAGQALSSQOE 416  
QY 314 VSOHLHAGNISEL-GINNNNR-----DLAFHLREVSDFRQSEPHSIFSLLEKAIK 366  
DB 417 LASSQKADKIQLEKELONNQKRSSELEFANMVRSLTATLNS--NSSTELKOK-- 471  
QY 367 WGYLSPELLREPMSEONCD-ALSTIFNAGLHMLDVLLEPVSTPVGESPTQOAPK 425  
DB 472 -----LETLEKELQARQOTKALTEELINVLTS-----LAKEQQTQIQNLQTIQYM 520  
QY 426 SVSDPRSRVEEHSVOTSEVDTQSKODKPOSSATSAL 461  
DB 521 EVEKEKVE-----LVKVLQQAQSSSSAEAL 549  
  
RESULT 15  
Q01514  
ID 01514 PRELIMINARY; PRT; 1978 AA.  
AC 01514;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE TRIP230.  
GN TRIP230.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97404344; PubMed=9256431;  
RA Chang K.H., Chen Y., Chen T.T., Chou W.H., Chen P.L., Ma Y.T.,  
RA Yang-Feng T.L., Deng L., Tsai M.J., O'Malley B.W., Lee W.H.;  
RT "A thyroid hormone receptor coactivator negatively regulated by the  
RT retinoblastoma protein.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:9040-9045(1997).  
DR EMBL: AF007217; AAD09135.1; -.  
SQ SEQUENCE 1978 AA; 227514 MW; AACBAEBA5C75FD7E CRC64;  
  
Query Match 5.5%; Score 128; DB 4; Length 1978;  
Best Local Similarity 21.0%; Pred. No. 9.4;  
Matches 126; Conservative 77; Mismatches 191; Indels 206; Gaps 26;  
  
QY 4 SKQIQEQLSPSLSDSICGVYIKL-----EKS-----ARRPLRNFNVAQ 43  
DB 397 AENEIMKRLSLNODNSLAEDNLIKMKRIEVLKREKESILSOKKEELQMSLKLNNYEYVIAK 456

```

QY 44 -TALRKLS-----ONPSADERDALQEA 64
Db 457 STATROISLSELDLRLNLEAKQELNOSISKEFTLIAIEELDRONQATKHMILIKA 516
QY 65 CLNKKILSDSLYEQFSKTTTRDIEL-----ISWFAAQFLIDTTLESANSL-- 111
Db 517 QLSKQONEGDSIISKLODINDEKRRVHOLEDDKMDITKELDVQFVLLIQSEVALNDLHL 576
QY 112 --EMLADLSEKHMHDHNPVLPVETLKSDDDKGERQADAKAFOLVGDSESSITLYA 169
Db 577 TKOKLEDEKVENLVQNLN-----KSQESNWSIQENLELKEHIRO-----NEEELSRIRN 625
QY 170 PVLQ-LPLVGEVTFPFQSAERKEISOLKSMLT---TTVAOERFAIQFKMENAK---- 220
Db 626 ELMQSLNODSNSNFKDTLKEREAVERNLKONLSELEQLNENLKVAFDVKMENKLYLA 685
QY 221 -----RCVTQDLRLS---ALVSTKCHSLGSSQSTNFGFAKSL----TRVENALVH 263
Db 686 CEDVRHOLEECLAGNNQLSLEKNTIVETLMEKEGEIAPELCWAOKRLLLEANKYEKTIEE 745
QY 264 LSG-----IKLAPKAERKTVQEVAVESSVSEGELPBMDTKHIERIPMA 307
Db 746 LSNARNLNTSALOHEHEHLIKLQOKR---DMEIAELKRN---IEQMDTDHKEETKQDL 796
QY 308 SEOAQTVSOHLHAGNLSELGNLNNMRDLAFHLLREVSDYFROSEPHSPISFLEK---A 364
Db 797 SSSLEQOKO-----LTQLIN---KKEIFIEKLKERSKLOEE-----LDKYSOA 837
QY 365 IRNGYLSLPPELLRE-----MMSEONGDALSTIFNAGLNLHDQVL--LPEVSTPT 412
Db 838 LRKN-----EILRQITEEKDRSLGSMKEEN-----NHLQEELERLREOSRT 879
QY 413 VGIESPQT---PQAKPSVSDPRSVVEHV-----SQTSFVDTQSKQOQKPO 454
Db 880 APVADPPTLDSVTELASVSQLNTIKHEHEEIKHHQKIIEDONQSKMOLLQSIQEOQKE 939

```

Search completed: March 2, 2003, 04:11:38  
Job time : 85 secs

